

Erretrobirus endogenoen detekzioa eta karakterizazioa behietan eta beste hainbat mamaliotan

Egilea: **KOLDO GARCIA ETXEBARRIA**

Urtea: 2010

Zuzendariak: **BEGO JUGO ORRANTIA**

Unibertsitatea: **UPV/EHU**

ISBN: 978-84-8438-587-5

Hitzaurrea

Euskara ama-hizkuntza dut. Ikasketak (Arizko Ikastola, Urbi Institutua, Euskal Herriko Unibertsitatea) euskaraz egin nituen, karrera eta doktoradutza-programa euskara hutsean egiteko aukera izan ez banuen ere. Genetikako laborategian sartu nintzenean lan-taldearen lan-hizkuntza euskara zen. Gauzarik naturalena eta normalena tesia euskaraz egitea zen eta bai ikertzeko nahikotasuna eta tesia euskaraz idatzi eta defendatu nituen. Beste edozein hizkuntz normal bat bezala.

Tesiaren erronka hirukoitza zen: genoma-osoko ikuspegia landu kasu handirik egiten ez zaien osagai genetikoei buruz, soilik lanabes bionformatikoak erabilia egin eta euskaraz idatzi eta defendatu. Erronkak baina ez zailtasunak, horiek beste leku batzuetatik etorri ziren. Tesia hiru arlo horietan berritzailea zenez dena egiteke zegoen eta erronka horiei gogotsu ekin nien. Gustora egiten den lana. Euskararen arloan hitzak asmatu behar izan genituen, *query* itzultzeko amu-sekuentzia erabiltzea nire gogokoena izan zen. Tesiaren hasieran euskara – ingelesa – gaztelera hiztegitxo bat aurkituko duzu horiek biltzen dituenak.

Tesiaren helburua zientifikoa bazen ere, euskaraz egin izanak hizkuntzarekiko erantzunkizun eta konpromezu bat ekarri zuen. Bilatuta izan ez bazen ere hizkuntza hauspotu, indartu eta zabaldu zen; eta euskara ere goi-mailako hizkuntza dela, berriro ere, frogatu. Bai, euskaraz gai abstraktueta eta orain arte ez jakindakoetaz aritu gaitezke beste edozein hizkuntzetan bezala. Ezagumendu berria euskaraz sortu daiteke, mugarik gabe. Mugatu gabe. Konplexurik gabe.

Eskuetan duzun tesi hau ez da bakarrik euskararen hauspotze, indartze eta zabaltzeari egindako ekarpen bat. Ezta beranduago Euskal Herriko Unibertsitateak eta Euskaltzaindiak Zientzia esperimental arloan II. Euskarazko Tesien Koldo Mitxelena Sariarekin ohoratutako tesi bat. Ezta ere euskara teknikoarentzat ondare soila. Baita ere tesi hau gurasoei eta arbasoei omenaldia eta esker ona da, euskara gaur egun arte transmititzeagatik eta oparitzeagatik.

Basaurin, 2016ko apirilaren 27an.

Koldo Garcia

eman ta zabal zazu



Universidad
del País Vasco

Euskal Herriko
Unibertsitatea

Erretrobirus endogenoen detekzioa eta karakterizazioa behietan eta beste hainbat mamaliotan

Koldo Garcia Etxebarria
2010

Euskal Herriko Unibertsitateko Genetika, Antropologia Fisikoa eta Animali Fisiologia Sailean egindako tesi hau, UPV/EHUK emandako EHU06/107 proiektuarekin, UPV/EHU eta enpresa arteko kontratuarekin eta UPV/EHUK emandako GIU07/62 diru-laguntzarekin finantzatu da.

Egileak 2006 urtean UPV/EHUko Euskara Errektoreordetzako Euskarazko tesiak sustatzeko beka eta 2006-2009 urteen bitartean UPV/EHUko Euskara Errektoreordetzako Euskarazko tesiak sustatzeko diru-laguntza jaso ditu.

Atxitxe eta Amamari
(Luis Etxebarria eta Aurelia Garmendiari)
Gure bihotzetan harro zaudetelako

A Amama Dámasa
(Dámasa García)
por llegar a vivirlo

*“Eskerrik asko bizitzari, horrenbeste eman didanak
irria eman dit eta negarra eman dit.”*
(Violeta Parra)

Bere buruaz beste egin zuen pertsona batek hitz eder hauek idatzi izanak betidanik deigarria egin zait. Eta egon daitekeen kontraesanik handiena. Baina azken finean bizitza hori da, kontraesanez betetako bide zoragarri bat. Eta eskerrak ematerakoan tesia izan den etapa honetan lagundu nautenei, bizitzan lagundu nautenei eta, azken finean, bizitzari eskerrak eman behar dizkiet. Bizitza eta lana banatzeak zentzurik ez daukalako, garenak gure bizitzako esparru guztiak bustitzen dituelako. Bide honetan, ezer ulertu gabe bazen ere, pausu bakoitzean arduratu direnei. Asko zarete abentura hau nola zihoan galdetu didazuenak eta bidaia atseginagoa egiten saiatu zaretenak. Ez daukat hitzik eman didazuen adore guztia eskertzeko. Badakit jendea ahaztu egingo zaidala, hala ere...

Lehenik eta behin Begori eman behar dizkiot. Zure konfiantzarik gabe guzti hau ezinezkoa izango baitzen. Nigan sinistu zenuen inork egiten ez zuen lan modu bat aurrera eramateko. Asko pasa dugu, lan asko eta txukuna egin dugu, asko ikasi dut eta, gainera, bakoitzaren bizitza aldatu da, eta nola gainera! Baina beti gauza arraroak egiten dituzten horiek izango gara.

MHC-koak direnei edo MHC-tik pasa direnei. Bereziki Amaia, hasieratik batera egon gara eta urte hauetan denetarik pasa eta partekatu dugu: lan eremua, ordenagailuak, kongresuak, kafeak, zurrumurruak, lanak, oroigarriak, azken orduko papeleo eta posterrak, etxeetara bisitak... eskerrik asko momentu guztiengatik. Maialeni, ez gara bakarrik laborategian batera egon, azken bi urteetan ingelesean ere, gaindosiarekin bukatuko duzu, eta gainera tesi honek duen lan esperimental bakarra zuk egin duzu, meritu handia daukazu. Ibairi, zure onddoen zuhaitz filogenetikoekin baso bat landatu (edo suntsitu) badaiteke ere, zurekin lan egiterakoan asko ikasi dut, baina, batez ere, oso atseginak izan dira zurekin partekatutako tarte guztiak. Inakiri, bakarrik *frango churrasco* agatik eskerrak merezi dituzu, baina zure sugandilek (eta eman didaten lanek) ere aipamena merezi dute eta, nola ez, pasatutako abentura guztiak, Donostian galduta barne. Eta urte hauetan MHCtik pasa diren gainontzekoei. Beti MHC izango gara eta Rupert gure profeta.

Ezin naiteke ahaztu *laboko* gainontzeko jendeaz. Eskerrik asko ni bezalako arraro baten lana pixka bat ezberdina dela ulertzeagatik eta barkatu sortutako arazoak, gaizki-ulertuak eta emandako *txarleta* ulergaitz horiek. Eta ordenagailuei buruz zeozer galdetu didatenei, nigan konfiantza hori edukitzea eskertzen dizuet. Bereziki Iratxeri, ez bakarrik egin dizkidazun milaka galderengatik, batez ere Bartzelonan primeran pasa genuelako, eta, gogoratu, ez hartu ezer lurretik. Eta baita kafetik pasa den zerrenda bukaezin horri ere eskerrak eman behar dizkiot, orduak eta orduak pasa baititugu gure kontuak partekatzen. Eta Antroko jendeari, Montseri eta Ikerri, hitz egiten pasa ditugun orduengatik. Guztiontzat bizkotxo!

Tesi hau aurrera eramatea posible egin duten beste gauza batzuk ere badaude. Derrigorrez abusatu dudan ordenagailu guztiei eskerrak eman behar dizkiet, euren sufrimendu gaitasunagatik ez balitz ez nituzke orri hauek beteko. Bereziki *Sagartxuri*, nire *Mavari*. Eta *Forfiri*, nire kotxea, urte hauetan egindako zerbitzu ikaragarriagatik, ez bakarrik niri, baita hemen agertzen den jende askori ere. Eta lan guztia, bidaiak eta idazketa lagundu duten soinu-bandaren partaideei. Eskerrik asko Amparo Sanchez-i “Seguiré caminando”-gatik, Quique González-i “Ajuste de cuentas”-gatik, Lichis-i “Ni jaulas, ni peceras”-gatik, Wilco-ri “Kicking Television”-gatik eta Jack Johnson-i “En concert”-gatik, bizitza zuzenean soinu hobea duelako. Coen Anaiei eta Pixarrekoie euren filmeengatik.

Euren sorbaldetan bidaitu dudan erraldoi guztiei.

Egunero ez egonda ere arduratu diren guztiei: Juanjo eta Nieves; Karol eta Beni; Euskarabilako jendea; Euskal blogosferakoak eta gainontzeko ziberrak.

Aitorri eta Bergoiri, urte hauetan egindako bazkariak, bidalitako mezuak eta emailak eta bisitak eskertzen dizkizuet, karrerako momentu onenak gogoratuz, bazkarian egindako telefono deiak boikotatuz eta kamisetak prestatuz.

Leireri, lankide izatetik laguna izatera pasa zarela. Urte hauetan pazientzia handia izan duzu nirekin eta nire izaerarekin, gure kafe amaigabe horietan. Badakit askotan txantxekin pasa naizela, baina badakizu zu bezalako gotika zapuztu bati adarra jotzea gustuko dudala eta beti intentzio

onekin izan dela. Eta baita zure etxeko eta bizitzako ateak zabaltzeagatik (eta hedapenez, Ikerri eta Neguri). Evari, printzesari, zuri ere gauza berberagatik, Leireren bizitza lotsa gabe asaltatu dugu eta oraindik ez dut ulertzen nola ez gaituen pikutara bidali. Eskerrik asko biei bazkariengatik, galdetzeagatik, hor egoteagatik. Ez dakit zer egingo duzuen ni gabe. Eta Arantzari, Leireren etxea okupatzen dugunetan, ez dezatela zure irria itzali.

Senitartekoei. Ez duzue ezer ulertuko baina harro zaudete. Hori asko da. Beti galdetzen, beti arduratuta, beti presioa kentzen saiatzen. Eskerrik asko Gabonak mantentzeagatik. Eskerrik asko Amama Dámasa, Tía Isabel, Tío Agustín, Ana, Aitor, Uxue (familiako fitxaketa berria), Gabi, Maider; Izeko Bego, Osaba Elías, Amagoia, Osabita, Izeko Edurne, Unai; eta Eneritzi, eskerrik asko danagatik, eta barkatu jostailuak jaso ez izana, zeu baino beranduagoa Doktore izango naiz, baina zeu beti *doktora muñe* izango zara eta ni *doktor txilin*.

Maiderrri. Bide honetan ezinbesteko euskarria izan zara. Beti laguntzeko, entzuteko, jasateko prest egon zara. Eskerrik asko inork ulertzen ez zuena ulertzeagatik, inork ulertzen ez dituen elkarrizketak edukitzeagatik eta, batez ere, nire Laguna izateagatik. *Llu nou*.

Aitari eta Amari.

Eskerrik asko 1982ko Otsail horretan salto ez egiteagatik eta ahal izan zenuten guztia egiteagatik. Eta ordutik egindako guztia. Zuen urrezko haurrak tesi hau idatzi dizue.

Eskerrik asko, bihotzez.

*“Hau da niretzat nabi nuen bizitza,
baina ez da niretzat nabi zenuen bizitza.”*
(Quique González)

Koldo

*“Gracias a la vida que me ha dado tanto
me ha dado la risa y me ha dado el llanto”.*
(Violeta Parra)

Siempre me ha parecido llamativo que una persona que se suicidó escribiese estas palabras tan bellas. Y la mayor contradicción que puede haber. Pero al final la vida es eso, un camino maravilloso lleno de contradicciones. Y al dar las gracias se los tengo que dar a los que me han ayudado en esta etapa que ha sido la tesis, a los que me han ayudado en la vida y, a la postre, a la vida. Porque separar vida y trabajo no tiene sentido, lo que somos empapa todos los ámbitos de nuestra vida. A los que en este camino, aunque no entendiesen nada, a cada paso se han preocupado. Sois muchos los que me habéis preguntado cómo iba esta aventura y los que habéis intentado hacerme este viaje más agradable. No tengo palabras para agradecerlos todo el aliento que me habéis dado. Sé que se me olvidará gente, sin embargo,...

En primer lugar tengo que agradecerlo a Bego. Ya que sin tu confianza todo esto no hubiese sido posible. Creíste en mí para sacar una forma de trabajo que no hacía nadie. Hemos pasado muchas cosas, hemos hecho mucho trabajo y bien hecho, he aprendido mucho y, no solo eso sino que, la vida de cada uno ha cambiado, y ¡cómo además! Pero siempre seremos esos que hacen cosas raras.

A los que son de MHC o los que han pasado por MHC. Especialmente a Amaia, hemos estado juntos desde el principio y en estos años hemos pasado y compartido de todo: zona de trabajo, ordenadores, congresos, cafés, cotilleos, trabajos, recuerdos, papeleos y pósteres de último momento, visitas a domicilio,... muchas gracias por todos los momentos. A Maialen, que no solo hemos estado juntos en el laboratorio, sino también los dos últimos años en inglés, terminarás con sobredosis, y además el único trabajo experimental que tiene esta tesis la has hecho tú, tienen un gran mérito. A Ibai, aunque con los árboles filogenéticos de tus hongos se puede plantar un bosque (o destruirlo), he aprendido mucho trabajando contigo, pero sobre todo, todos los momentos que he compartido contigo han sido muy agradables. A Iñaki, solo ya por el *frango churrasco* mereces las gracias, pero también merecen ser mencionadas tus lagartijas (y el trabajo que me han dado) y, como no, todas las aventuras que hemos pasado, incluido estar perdido en Donostia. Y al resto que ha pasado por MHC. Siempre seremos MHC y Rupert nuestro profeta.

No me puedo olvidar del resto de la gente del *labo*. Gracias por entender que el trabajo de un raro como yo es distinto y os pido perdón por los problemas ocasionados, los malentendidos y las *txarletas* incomprensibles que he dado. Y a los que me habéis preguntado algo sobre ordenadores, os agradezco esa confianza que depositasteis en mí. Especialmente a Iratxe, no solo por las miles de preguntas que me has hecho, sobre todo porque lo pasamos estupendamente en Barcelona, y, recuerda, no cojas nada del suelo. También tengo que darles las gracias a esa lista interminable de personas que han pasado por los cafés, ya que hemos pasado horas y horas compartiendo nuestras cosas. Y a la gente de Antro por las horas que hemos pasado hablando. ¡Para todos bizcocho!

También hay otros tipos de cosas que han hecho posible hacer esta tesis. A la fuerza tengo que dar las gracias a todos los ordenadores de los que abusé, si no fuese por su capacidad de sufrimiento no hubiese rellenado estas páginas. Especialmente a *Sagartxu*, mi *Mac*. Y a *Forfi*, mi coche, por el increíble servicio prestado estos años, no solo a mí, sino también a un montón de gente que aparece aquí. A los participantes de la banda sonora que han acompañado todo el trabajo, los viajes y la escritura. Gracias a Amparo Sanchez por “Seguiré caminando”, a Quique González por “Ajuste de cuentas”, a Lichis por “Ni jaulas, ni peceras”, a Wilco por “Kicking Television” y a Jack Johnson por “En concert”, la vida en directo suena mucho mejor. A los Hermanos Coen y a los de Pixar por sus películas.

A todos los gigantes en cuyos hombros he viajado.

A los que, aunque no han estado día a día, los que se has preocupado: Juanjo y Nieves; Karol y Beni; la gente de Euskarabila; los de la blogosfera vasca y demás cibers.

A Aitor y Bergoi, os agradezco las comidas hechas estos años, los mensajes e emails mandados y las visitas, recordando los mejores momentos de la carrera, boicoteando las llamadas hechas en las comidas y preparando camisetas.

A Leire, porque has pasado de ser una compañera a una amiga. Estos años has tenido una gran paciencia conmigo y con mi forma de ser, en esos nuestros cafés interminables. Sé que muchas veces me he pasado con las bromas pero sabes que me encanta tomar el pelo a una gótica frustrada como tú y que siempre lo he hecho con buena intención. Y también por abrirme las puertas de tu casa y de tu vida (y por extensión, a Iker y Negu). A Eva, printzesa, a ti también por lo mismo, hemos asaltado la vida de Leire sin vergüenza alguna y aún no entiendo cómo no nos ha mandado a freír espárragos. Gracias a las dos por las comidas, por preguntar, por estar ahí. No sé que haréis sin mí. Y a Arantza, por cuando okupamos la casa de Leire, que no apaguen tu sonrisa.

A la familia. No entenderéis nada pero estáis orgullosos. Eso es mucho. Siempre preguntado, siempre preocupados, siempre intentando quitar presión. Gracias por mantener las Navidades. Gracias a Amama Dámasa, Tía Isabel, Tío Agustín, Ana, Aitor, Uxue (familiako fitxaketa berria), Gabi, Maider; Izeko Bego, Osaba Elías, Amagoia, Osabita, Izeko Edurne, Unai; y a Eneritz, gracias por todo, y perdona no haber recogido los juguetes, seré Doctor más tarde que tú, pero tú siempre serás *doktora muñe* y yo *doktor txilin*.

A Maider. Has si un soporte imprescindible en este camino. Siempre has estado dispuesta a ayudar, a escuchar, a aguantar. Gracias por entender lo que nadie entiende, por tener esas conversaciones que nadie entiende y, sobre todo, por ser mi Amiga. *Llu nou*.

A Aita y Ama.

Muchas gracias por no haber saltado en aquel febrero de 1982 y por hacer todo lo que podíais Y todo lo hecho desde entonces. Vuestro niño de oro os ha escrito esta tesis.

Muchas gracias, de corazón.

*“Esta es la vida que yo quería para mí,
pero no es la vida que tú querías para mí”*
(Quique González)

Koldo

Aurkibidea

Aurkibidea.....	V
Taulen aurkibidea.....	VII
Irudien aurkibidea.....	IX
Laburdurak.....	XI
Hiztegi laburra.....	XIII
Nomenklatura.....	XIII
Laburpena.....	XV
Summary.....	XVII
Resumen.....	XIX
I. Sarrera.....	3
I.1. ERVen egitura.....	3
I.2. ERVen sailkapena eta banaketa.....	5
I.3. ERVen in silico azterketa.....	8
I.4. ERVen dinamika ebolutiboa.....	12
I.5. ERVen adierazpena eta ostalarian duten eragina.....	18
I.6. Aztertutako taxoiak eta espezieak.....	22
I.7. Lanaren planteamendua.....	24
I.8. Bibliografia.....	26
II. Genome-wide detection and characterization of endogenous retroviruses in Bos taurus.....	31
II.1. Abstract.....	33
II.2. Introduction.....	33
II.3. Materials and methods.....	34
II.4. Results.....	36
II.5. Discussion.....	44
II.6. References.....	46
II.7. Appendix.....	48
III. Zaldiaren erretrobirus endogenoen (ERV) detekzioa eta sailkapena.....	51
III.1. Sarrera.....	53
III.2. Material eta metodoak.....	54
III.3. Emaizak.....	56
III.4. Eztabaida.....	63
III.5. Bibliografia.....	65
III.6. Eranskina.....	68
IV. Mamalioen erretrobirus endogenoen (ERV) dinamika ebolutiboa.....	73
IV.1. Sarrera.....	75
IV.2. Material eta metodoak.....	76
IV.3. Emaizak.....	79
IV.4. Eztabaida.....	102
IV.5. Bibliografia.....	106
IV.6. Eranskina.....	109
V. Genes around bovine endogenous retroviruses (BoERVs) are related to host defense and histone functions.....	133
V.1. Introduction.....	135
V.2. Materials and methods.....	136
V.3. Results.....	137
V.4. Discussion.....	142
V.5. References.....	146
V.6. Appendix.....	148
VI. Behiaren erretrobirus endogenoen (BoERV) adierazpena behiaren (Bos taurus) EST-etan .	153
VI.1. Sarrera.....	155
VI.2. Materialak eta metodoak.....	156
VI.3. Emaizak.....	157
VI.4. Eztabaida.....	162
VI.5. Bibliografia.....	165
VI.6. Eranskina.....	167
VII. Eztabaida orokorra.....	173
Bibliografia.....	180
VIII. Ondorioak.....	183

Taulen aurkibidea

I-1. taula: ERVen in silico bilaketak egindako genomak.	10
I-2. taula: HERV familien kopia-kopuruaren emendioaren mekanismo nagusiak.	15
I-3. taula: Mikromatritzeen bidez egindako ERVen adierazpenari buruzko lanen laburpena.	21
Table II-1: Previously characterized retroviruses from a variety of species used as query.	35
Table II-2: ERV elements detected by each method.	37
Table II-3: Structure of ERV elements detected by each method.	38
Table II-4: Properties of BoERV families characterized in this work.	40
III-1. taula: Bilaketa genomikoa burutzeko erabilitako erretrobirusak, amu-sekuentzia posizioa eta sekuentzien arteko distantzia.	55
III-2. taula: Zaldiriaren genoma-bertsio bakoitzean detektatutako elementu-kopuruaren arteko ezberdintasuna.	57
III-3. taula: Metodo bakoitzak detektatutako elementuak EquCab2 genoma-bertsioan.	58
III-4. taula: Metodo bakoitzak detektatutako elementuen egitura EquCab2 genomaren bertsioan.	59
III-5. taula: Detektatutako EqERV familiak.	61
III-6. taula: Behiaren eta zaldiriaren ERVen detekzioaren konparazioa.	63
IV-1. taula: Aztertutako genomaren laburpena.	76
IV-2. taula: Bilaketa genomikoa burutzeko erabilitako erretrobirusak, amu-sekuentzia posizioa eta sekuentzien arteko distantzia.	77
IV-3. taula: Aztertutako ERVen ezaugarriak espezie eta metafamilia bakoitzean.	80
IV-4. taula: Errekonbinazio intergeniko gertaren laburpena, aztertutako filogenia ezberdinen geneetan.	89
IV-5. taula: ERVen hautespenaren estimazioa aztertutako espezie, metafamilia eta geneetan. Adar guztietan ω berdina (M0 eredua) izatea eta barneko eta kanpoko adarrek ω ezberdina izatearen artean eredu probableena aztertzen duen analisiaren laburpena.	90
IV-6. taula: ERVen errekonbinazio intragenikoaren detekzioa aztertutako espezie, metafamilia eta geneetan.	93
IV-7. taula: ERVen hautespenaren estimazioa aztertutako espezie, metafamilia eta geneetan. ω bakarria egotearen (M0 eredua) eta 3 ω ezberdin egotearen (M3 eredua) artean eredu probableena estimatzen duen analisiaren laburpena.	97
IV-8. taula: ERVen hautespenaren estimazioa aztertutako espezie, metafamilia eta geneetan. Eredu ia neutralaren (M1 eredua) eta hautespen positiboaren (M2 eredua) artean eredu probableena estimatzen duen analisiaren laburpena.	99
Table V-1: Features of used datasets.	136
Table V-2: Overview of gene and BoERVs associations in each range and BoERV data-set.	137
Table V-3: Features of all genes near BoERVs, in each range and BoERV data-set.	139
Table V-4: Overrepresented Gene Ontology terms in each range and BoERV data-set.	140
Table V-5: Features of BoERVs around genes that are in overrepresented GO terms, in each range and BoERV data-set.	141
Table V-6: Features of genes related with overrepresented GO terms in each range and BoERV data-set.	142
VI-1. taula: Adierazitako diren behiaren BoERVen inguruan dauden gene-kopurua.	161
VI-2. taula: Gene Ontology termino gainerrepresentatuak eta erlazioatutako ERV eta geneak.	162
VII-1. taula: ERV eta genomaren arteko erlazioen laburpena.	178

Irudien aurkibidea

I-1. irudia: Elementu transposakorren sailkapena eta eretroelementuen egitura.	3
I-2. irudia: Erretrobirusen oinarritzko egitura.	3
I-3. irudia: ERVen banaketa hainbat espezieetan.	6
I-4. irudia: LTRdun elementuak detektatzeko LTR_STRUC-ek bilatzen dituen egiturak (McCarthy & McDonald, 2003).	9
I-5. irudia: ERVen bilaketa egiteko Retroector©-ek erabiltzen dituen motiboak (Sperber et al., 2007).	9
I-6. irudia: ERVen bizi-zikloa.	12
I-7. irudia: Erretrobirusen aurkako defentsak.	13
I-8. irudia: KoRVren agerpena koalaren hainbat populaziotan (Tarlinton et al., 2006).	17
I-9. irudia: enJSRVen eta ardiaren historia ebolutiboa. (Arnaud et al., 2008).	17
I-10. irudia: Elementu transposakorrak eta genoma-ingurune elkarrekintzak.	18
Figure II-1: Diagram representing the number of detected elements (circle size) and common elements (overlapping area).	38
Figure II-2: RT region based phylogenetic tree of BoERVs.	39
Figure II-3: Partial aminoacid sequence of the RT region of the representative ERVs from each putative family.	42
Figure II-4: RT region based unrooted phylogenetic tree of ERVs from different species.	43
III-1. irudia: Aztertutako zaldiaren bi genoma-bertsioetan detektatutako elementu-kopurua (zirkuluen tamainarekiko) eta elementu komunak (gainjartzen den azalera) irudikatzen dituen Venn diagrama.	56
III-2. irudia: RT eskualdean oinarritutako EeERV-en zuhaitz filogenetikoa.	60
III-3. irudia: Espezie ezberdinetako ERVen RT eskualdean oinarritutako errota gabeko zuhaitz filogenetikoa.	62
IV-1. irudia: Aukeratutako gizakiaren ERVen gag+pol+env geneen gene anizkoitzeko zuhaitz filogenetikoa.	81
IV-2. irudia: LTRak detektatu zaien gizakien ERVen 5'LTR, parte zentrala eta 3'LTR zuhaitz filogenetikoak.	82
IV-3. irudia: Aukeratutako txinpantzearen ERVen gag+pol+env geneen gene anizkoitzeko zuhaitz filogenetikoa.	82
IV-4. irudia: Aukeratutako saguaren ERVen gag+pol+env geneen gene anizkoitzeko zuhaitz filogenetikoa.	83
IV-5. irudia: LTRak detektatu zaien saguen ERVen 5'LTR, parte zentrala eta 3'LTR zuhaitz filogenetikoak.	84
IV-6. irudia: Aukeratutako arratoiaren ERVen gag+pol+env geneen gene anizkoitzeko zuhaitz filogenetikoa.	84
IV-7. irudia: LTRak detektatu zaien arratoien ERVen 5'LTR, parte zentrala eta 3'LTR zuhaitz filogenetikoak.	86
IV-8. irudia: Aukeratutako txakurraren ERVen gag+pol+env geneen gene anizkoitzeko zuhaitz filogenetikoa.	87
IV-9. irudia: Aukeratutako behiaren ERVen gag+pol+env geneen gene anizkoitzeko zuhaitz filogenetikoa.	87
IV-10. irudia: LTRak detektatu zaien behien ERVen 5'LTR, parte zentrala eta 3'LTR zuhaitz filogenetikoak.	88
IV-11. irudia: Aukeratutako zaldiaren ERVen gag+pol+env geneen gene anizkoitzeko zuhaitz filogenetikoa.	88
Figure V-1: Insertion orientation of BoERVs in each range and data-set.	137
Figure V-2: Number of associations of BoERVs and genes in each dataset.	138
VI-1. irudia: gag, pol eta env ERVen geneen eta GAPDH genearen adierazpen erlatiboa aurkitutako ESTak ehun/organoka sailkatuta.	157
VI-2. irudia: gag genearen adierazpen-maila erlatiboa ((ESTak/ehunaren EST kopuru totala)*105) ehun eta familiak.	158
VI-3. irudia: pol genearen adierazpen maila erlatiboa ((ESTak/ehunaren EST kopuru totala)*105) ehun eta familiak.	159
VI-4. irudia: env genearen adierazpen maila erlatiboa ((ESTak/ehunaren EST kopuru totala)*105) ehun eta familiak.	159
VI-5. irudia: Adierazpena aztertu zaien ERVen familien zuhaitz filogenetikoa.	160
VI-6. irudia: BoERVen txertaketa-norabidea.	161
VII-1. irudia: Tesi-lan honetan egindako analisisien adierazpen eskematikoa.	175
VII-2. irudia: Erretrobirus endogenoek buruzko ikuspegi orokorra.	176

Laburdurak

BFV (*Bovine foamy virus*): behi apardunen birus.

BLV (*Bovine leukemia virus*): behien leuzemiaren birus.

BoERV (*bovine endogenous retrovirus*): behien erretrobirus endogeno.

CERV (*chimpanzee endogenous retrovirus*): txinpantzeen erretrobirus endogeno.

EIAV (*Equine Infectious Anemia Virus*): zaldien anemia infektziosoaren birus.

EqERV (*equine endogenous retrovirus*): zaldien erretrobirus endogeno.

ERV (*endogenous retrovirus*): erretrobirus endogeno.

EST (*expressed sequence tag*): adierazitako sekuentzien markatzaile.

FDR (*false discovery rate*): aurkikuntza faltsuen tasa

FeLV (*Feline leukemia virus*): felidoen leuzemia birus.

GALV (*Gibbon Ape leukemia virus*): Giboi tximinoaren leuzemiaren birus.

GO (*Gene Ontology*): gene-ontologia

HERV (*human endogenous retrovirus*): gizakien erretrobirus endogeno.

HIV (*Human immunodeficiency virus*): giza immunoeskasiaren birus.

HTLV (*Human T cell lymphotropic virus*): giza T zelula linfotrofikoen birus.

HSRV (*Human spuma retrovirus*): giza spumaretrobirus.

IGF2 (*Insulin-like growth factor 2*): insulina-antzerako hazkunde-faktore 2.

IRG (*Immunity related GTPases*): immunitatearekin erlazionatutako GTPasak.

JSRV (*Jaagsiekte Sheep Retrovirus*): ardien Jaagsiekte retrobirus.

LTR (*long terminal repeats*): muturreko errepika luzeak.

Mf: METAFAMILIA.

MHC (*Major Histocompatibility Complex*).

ML (*maximum likelihood*): egiantz handien.

MLV (*Murine leukemia virus*): murinoen leuzemiaren birus.

MmERV (*Mus musculus endogenous retrovirus*): saguen erretrobirus endogeno.

MMTV (*Mouse mammary tumor virus*): saguen bular-tumorearen birus.

MPMV (*Mason-Pfizer Monkey Virus*): Mason-Pfizer tximino birus.

NJ (*neighbor joining*): aldamenekoen elkartzea.

OERV (*ovine endogenous retrovirus*): ardien erretrobirus endogeno.

PAG (*Pregnancy-associated glycoproteins*): haurdunaldiarekin erlazionatutako glikoproteinak.

PBS (*primer binding-site*): hasle-loturaren gunek.

PERV (*porcine endogenous retrovirus*): txerrien erretrobirus endogeno.

PIC (*preintegration complex*): txertaketa aurreko konplexu.

PPT (*polypurine tract*): polipurina traktua.

RnERV (*Rattus norvegicus endogenous retrovirus*): arratoien erretrobirus endogeno.

RT (*reverse transcriptase*): *pol* geneko alderantzizko transkriptasa eskualdea.

TE (*transposable element*): elementu transposakor.

TM (*transmembrane*): *env* geneko transmintza eskualdea.

XIST (*X chromosome inactivation-specific transcript*): X kromosomaren inaktibazioaren transkripto espezifikoa.

XRV (*exogenous retrovirus*): erretrobirus exogeno.

Hiztegi laburra

Euskara	English	Castellano
Adierazitako sekuentzien markatzaile	Expressed sequence tag (EST)	
Akordio-sekuentzia	Consense sequence	Secuencia consenso
Amu-sekuentzia	Sequence to search	
Antzekotasun	Similarity	Similitud
Aurkikuntza faltsuen tasa	False discovery rate	
Bikote-ezabapena	Pairwise deletion	
Eboluzio-esparru	Evolutionary framework	
Egiantz handien	Likelihood	Máxima versimilitud
Egiantza-arrazoi test	Likelihood ratio test	Test de máxima verosimilitud
Elementu transposakor	Transposable element	Elemento transponible
Estaldura	Cover	
Etxekotze-prozesua	Domestication	Domesticación
GC-edukia	GC content	Contenido GC
Gene-konbertsioa	Gene conversion	Conversión génica
Genearen inpronta patroia	Gene's imprinting pattern	
Genoma-bertsio	Genome version	Versión de la secuenciación del genoma
GO prozesu biologikoa	GO biological process	
Hasle-loturaren gune	Primer binding site	
Insertzio-denbora	Insertion time	Tiempo de inserción
Inpronta-patroia	Imprinting pattern	
Markov-en ezkutuko ereduak	Hidden Markov Models	Modelos ocultos de Markov
Murritzte-faktore	Restriction factor	Factor de restricción
Muturreko errepika luze	Long terminal repeat	
Ondorengo probabilitate bayestarra	Bayesian posterior probability	Probabilidad posterior Bayesiana
p-distantzia	p-distance	
solo-LTRak	solo-LTRs	
Talde kanpoko	Outgroup	
Txertaketa aurreko konplexu	Preintegration complex	
Txertatze-norabidea	Insertion direction	Dirección de la inserción

Nomenklatura

Lan honetan nomenklatura bi mailatan ezarri da: bata espezie bateko ERVak orokorrean izendatzeko eta bestea ERV bakoitza izendatzeko.

- Beste espezieetan erabilitako nomenklatura jarraituz, lan honetan behiaren ERVak BoERV (*Bovine endogenous retrovirus*) izendatu dira eta zaldiarenak EqERV (*Equine endogenous retrovirus*).

- ERV indibidualak modu estandarrean izendatzeko proposamena egin bada ere (Blomberg et al., 2009), arrazoi praktikoak direla eta, lan honetan ERV indibidualak izendatzeko zein kromosometan detektatu den eta kromosoma horretako zenbatgarren ERVa den erabili dira, adibidez, chr1_1, lehengoko kromosoman lehenengo agertzen den ERV adierazteko.

Laburpena

Lan honetan hainbat mamalioen genoman detektatutako erretrobirus endogenoak (ERV) lanabes bioinformatikoen bidez aztertu dira. ERVak jatorri erretrobirala duten elementu genomikoak dira eta ornodunen genomen hein zabal batean detektu dira, batez ere mamalioetan. Hala ere elementu hauei buruzko ezagumendua mugatua da eta egindako azterketa genomikoak primate eta karraskarietara mugatu da.

Honela, ERVen identifikazioa behian eta zaldian burutu da, zeintzuen ERVak maila genomikoan aztertu ez diren bi taxoen ordezkariak diren (Ruminantia eta Equidae). ERVen dinamika ebolutiboa (euren proliferazioa eta sekuentzien aldaketen baimentzea) aztertu da ere bi espezie hauen ERVetan, eta emaitzak testuinguruan jartzeko, beste bost espezien ERVak, hala nola, gizakia, txinpantzea, sagua, arratoia eta txakurra, ere aztertu ziren. Azkenik, behiaren genoman, ERVen adierazpena EST datubaseak erabiliz aztertu da eta ERVetatik gertu zeuden geneak karakterizatu dira.

ERVen detekzioa burutzeko hiru metodo ezberdin erabili ziren: BLAST algoritmoan oinarritutako metodo bat eta beste bi LTR_STRUC eta Retrotector© programa espezifikoetan oinarrituta. Metodoen artean ERV gehien aurkitu zituen Retrotector© izan bazen ere, komunak ez ziren emaitza kopuru altua zela eta, ERVak detektatzeko programa bat baino gehiago erabiltzea gomendagarria dirudi. Detektatutako ERVen analisi filogenetikoek behian 24 balizko ERV familia zeudela azaldu zuten, horietatik 20 esperimentalki ez zeuden deskribatuak eta lan honetan lehen aldiz deskribatzen den BoERV1 familia ugariena izan zen. Zaldian, aldiz, 15 balizko familia deskribatu ziren, aurretik deskribatuta ez zeudenak. Aztertutako espezie hauen arteko ezberdintasunaren jatorria genoma bakoitzaren dinamikan, espezie bakoitzaren ezaugarrietan eta etxekotze-prozesuan egon daiteke.

ERVen dinamika ebolutiboari dagokionez, ERVen kopia-kopuruaren emendioa eta ERVen sekuentziaren aldaketen baimentzea inferitzeko, errekonbinazio intergenikoa, errekonbinazio intragenikoa eta hautespena aztertu ziren. Sekuentzietan aldaketak baimentzeari dagokionez, oro har, ERVen sekuentzietan hautespen purifikatzailea nagusia izan zen, hautespen neutralpean eta hautespen positiboan zeuden kodoiak detektatu baziren ere. Gainera, zaldiaren ERVetan izan ezik, espezie gehienek ERVetan errekonbinazio intrageniko gertaera nahiko detektatu ziren. Gure analisen arabera, behiaren eta zaldiaren ERVak, gizaki, txinpantze, sagu eta txakurran bezala, berrinfekzio bidez euren kopurua emendatu zuten. Salbuespen bakarra arratoiaren ERVak izan ziren, zeinetan emaitzek erretrotransposizioa kopuruaren emendio-mekanismo moduan iradokitzen duten.

Testuinguru genomikoari dagokionez, behiaren ERVetatik gertu zeuden geneak, zati handi batean, birusen aurreko defentsarekin eta histonekin erlazionatutako geneak izan ziren. Gainera, ERVetatik gertu zeuden gene proportzio bat pseudogeneak eta auresandako geneak izan ziren, duplikazio gertaeren arrastoak izan daitezkeenak.

Azkenik, behiaren ESTak aztertuz behiaren ehunetan ERVen adierazpena detektatu da. Adierazpen hau garapen-fase, ehunaren eta aztertutako gene erretrobiralaren arabera izan zen. Gehien adierazi zen gene erretrobirala *env* izan zen eta ERVen adierazpena batez ere enbrioi-ehunetan gertatu zen.

Hortaz, lan honetan aurkeztu diren emaitzek, adibidez ERVen txertaketa preferentziak eta ERVen adierazpen kontrolatua, genoma ostalariaren eta elementu genomiko hauen artean erlazio bizi eta konplexua dagoela aditzera ematen dute.

Summary

In the present work endogenous retroviruses (ERV) detected on genomes from different mammals have been analysed by bioinformatic tools. ERVs are genomic elements with retroviral origin and they have been detected in a wide range of vertebrates, especially in mammals. However, the knowledge about this elements is limited and the genomic analyses performed were limited to primates and rodents.

Thus, the identification of ERVs have been performed in cow and horses, which are representatives of two taxa (Ruminantia and Equidae) whose ERVs have not had analyses at genomic level. Moreover, the evolutionary dynamic (their proliferation and the permissiveness of changes in the sequence) of ERVs of this all species has been studied also, and to place in context ERVs from other species, such as human, chimpanzee, mouse, rat and dog, have been analysed. Finally, in the bovine genome, the expression of ERVs have been analysed using the EST databases, and the genes around ERVs have been characterized.

In order to detect ERVs three different methods were used: one method based on BLAST algorithm and other two based on LTR_STRUC and Retroector© specific programs. Although Retroector© detected more elements, due to the high number of results that were not share, it seemed advisable to use more than one program to detect ERVs. The phylogenetic analyses of detected ERVs showed that in the cow there were 24 putative ERV families, from which 20 had not been experimentally described and the BoERV1 family, which is described in this work for the first time, was the largest. In the horse, however, 15 families were described, which had not been described before. The origin of the differences between these studied species could be the dynamic of each genome, the features of each species and the domestication process.

As far as ERV's evolutionary dynamic is concerned, to infer the proliferation of ERVs and the permissiveness of changes in the sequence of ERVs, intergenic recombination, intragenic recombination and selection were analysed. Regarding the changes allowed in the sequence, in general, purifying selection was the principal in the sequence of the 3 retroviral genes, although codons under neutral selection and positive selection were detected also. Moreover, except in horse's ERVs, there were detected many intragenic recombination events. According to the results of our analyses, ERVs from cow and horse, like human, chimpanzee, mouse and dog, proliferated by reinfection. The only exception were the ERVs from rat, where the results suggested retrotransposition as proliferation-mechanism.

As regards the genomic context, the genes located near bovine ERVs were, largely, genes related to defence against virus and histones. Moreover, a proportion of genes near of ERVs were pseudogenes or predicted genes, which could be traces of duplication events.

Finally, by the analysis of bovine ESTs, the expression of ERVs in bovine tissues was detected. This expression depended on the development-phase, tissue and studied retroviral gen. *env* gene was the most expressed retroviral gene and the highest expression of ERVs happened in embryonic tissue.

Thus, the processes detected in the present work such as the insertion preferences of ERVs and the control of the expression of ERVs, showed a dynamic and complex relation between the host's genome and these genomic elements.

Resumen

En este trabajo se han analizado los retrovirus endógenos (ERV) detectados en los genomas de varios mamíferos mediante herramientas bioinformáticas. Los ERVs son elementos genómicos de origen retroviral y han sido descritos en un amplio rango de genomas de vertebrados, sobre todo en mamíferos. Sin embargo el conocimiento sobre estos elementos es limitado y los análisis genómicos realizados se han circunscrito a primates y roedores.

Así, se ha realizado la identificación de ERVs en vaca y yegua, que son representantes de dos taxones (Ruminantia y Equidae) cuyos ERVs no han sido analizados a nivel genómico. También se ha estudiado la dinámica evolutiva (su proliferación y la permisividad de cambios en su secuencia) de los ERVs de estas dos especies, y para poner los resultados en contexto, se han analizado también los ERVs de otras especies, tales como, humano, chimpancé, ratón, rata y perro. Por último, en el genoma bovino, se ha analizado la expresión de ERVs utilizando las bases de datos de ESTs y se han caracterizado los genes cercanos a los ERVs.

Para realizar la detección de los ERVs se utilizaron tres métodos distintos: un método basado en el algoritmo BLAST y otros dos basados en los programas específicos LTR_STRUC y Retroector©. Aunque Retroector© fue el que más ERVs detectó, debido al alto número de resultados que no fueron comunes parece recomendable utilizar más de un programa para detectar ERVs. Los análisis filogenéticos de los ERVs detectados mostraron que en la vaca hay 24 posibles familias de ERVs, de las cuales 20 no estaban experimentalmente descritas y que la familia BoERV1, que se describe por primera vez en este trabajo, fue la más numerosa. En la yegua, por su parte, se describieron 15 posibles familias, que no habían sido descritas anteriormente. El origen de las diferencias entre las especies estudiadas podría radicar en la dinámica de cada genoma, en las características de cada especie y en el proceso de domesticación.

En cuanto a la dinámica evolutiva de los ERVs, para inferir la proliferación de los ERVs y la permisividad de los cambios en la secuencia de los ERVs, se analizaron la recombinación intergénica, la recombinación intragénica y la selección. En lo que respecta a los cambios permitidos en las secuencias, en general, la selección purificadora fue la mayoritaria en las secuencias de los 3 genes retrovirales, aunque también se detectaron codones bajo selección neutral y selección positiva. Además, excepto en los ERVs de la yegua, en los ERVs de la mayoría de las especies se detectaron numerosos eventos de recombinación intragénica. Según los resultados de nuestros análisis, los ERVs de vaca y yegua, al igual que de humano, chimpancé, ratón y perro, proliferaron mediante reinfección. La única excepción fueron los ERVs de rata, donde los resultados sugieren la retrotransposición como mecanismo de proliferación.

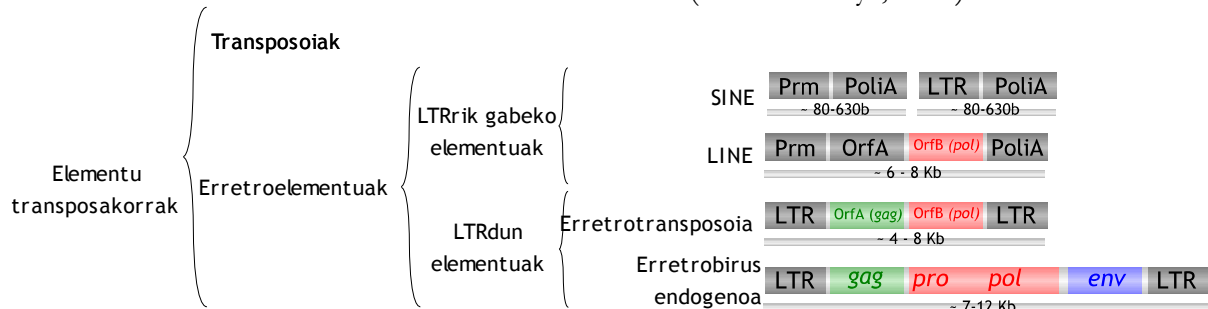
En cuanto al contexto genómico, los genes localizados cerca de los ERVs bovinos fueron, en gran parte, genes relacionados con la defensa ante virus y con histonas. Además, una proporción de los genes cercanos a los ERVs fueron pseudogenes o genes predichos, que pudieran ser rastros de eventos de duplicación.

Por último, mediante el análisis de ESTs bovinos se detectó la expresión de los ERVs en varios tejidos bovinos. Esta expresión dependió de la fase de desarrollo, del tejido y del gen retroviral examinado. El gen retroviral que más se expresó fue *env* y la expresión de los ERVs ocurrió sobre todo en el tejido embrionario.

Por tanto, los procesos detectados en este trabajo, tales como las preferencias de inserción de los ERVs y la expresión controlada de los ERVs, muestran que existe una relación dinámica y compleja entre el genoma huésped y estos elementos genómicos.

I. Sarrera

Erretrobirus endogenoak (ERV, *endogenous retrovirus* ingelesez) organismo eukariotoen genomatan agertzen diren elementuak dira. Erretrobirus exogenoek (XRV, *exogenous retrovirus*) hozi-lerroko zelulak infektatzen dituzte eta haien genomatan txertatzen dira. Hala, belaunaldiz belaunaldi transmititzen dira. ERV bihurtu direla esaten da orduan (Boeke & Stoye, 1997).



I-1. irudia: Elementu transposakorrak sailkapena eta erretroelementuen egitura.

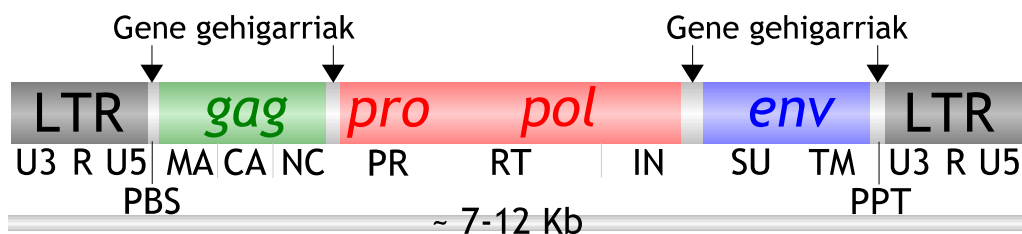
(Bannert & Kurth, 2004)-tik moldatua.

ERVak elementu transposakorrak (TE, *transposable elements*) deritzen elementu genomikoen barruan sailkatzen dira (I-1. irudia). Mamalioen genomaren ia erdia elementu transposakorrez osotuta dago. Elementu transposakorrak artean, bi talde bereizten dira: transposoiak eta erretroelementuak. Azken horiek, halaber, bi klase nagusitan banatzen dira: LTRrik (*long terminal repeats*) gabeko elementuak eta LTRdun elementuak. SINEak (*short interspersed elements*) eta LINEak (*long interspersed elements*) LTRrik gabeko klaseko kideak dira eta kopia kopuru handietan agertzen dira mamalioen hozi-lerroan; SINEek ez dute proteinarik kodetzeko gaitasunik eta LINEak behar dituzte amplifikatzeko. LTRdun klasearen barruan erretrotransposoiak, ERVak eta ERVdun jatorria duten elementu errepikakorrak (LTR bakartiak edo solo-LTRak, adibidez) daude (Bannert & Kurth, 2004).

I.1. ERVen egitura

Erretrobirus endogenoek (ERV) erretrobirus exogenoen (XRV) antzeko egitura dute: gutxienez *gag*, *pro-pol* eta *env* geneak eta alboetan LTRak azaltzen dituzte (I-2. irudia). Erretrobirus batzuetan, gene hauetaz gain, beste gene osagarri batzuk ere ager daitezke; horiek agertzen badira, erretrobirusak konplexutzat jotzen dira eta agertzen ez badira, sinpletzat (Vogt, 1997; Griffiths, 2001). Atal honetan azalduko den erretrobirusen genomaren informazio orokorra nagusiki Vogt-ek (1997) "Retrovirus" libururako egindako kapitulutik hartu da.

Geneez eta LTRez gain, beste bi gune garrantzitsu aurki daitezke erretrobirusen egiturari: hasle-loturaren gunea (PBS, *primer-binding site* ingelesez) eta polipurina traktua (PPT, *polypurine tract*). PBSa erretrobirusen alderantzizko transkripzioa hasteko gunea da, zeinari tRNA espezifiko bat lotzen zaion erretrobirusaren DNAREN (-) harizpiaren transkripzioa hasteko. PPTa erretrobirusaren DNAREN (+) harizpiaren haslea sintetizatzen duen eskualdea da (Vogt, 1997).



I-2. irudia: Erretrobirusen oinarriko egitura.

LTR, *long terminal repeats*, U3, R eta U5 eskualdeak. *gag* genea, matrizea (MA), kapsida (CA) eta nukleokapsida (NC) proteinak. *pro* genea, proteasa (PR) proteinak. *pol* genea, alderantzizko transkripataza (RT) eta integrasa (IN) entzimak. *env* genea, gainazal (SU) eta transmintza (TM) glikoproteinak. PBS, *protein binding site*, (-) harizpiaren sintesia hasteko haslea lotzen den eskualdea. PPT *polypurine tract*, (+) harizpiaren haslea sintetizatzen duen eskualdea.

1.1.1. *gag* genea

gag geneak birusaren barne-egiturako proteina bat kodetzen du, Gag proteina, alegia (jatorrizko *group-specific antigen* izenetik, gaitasun antigenikoak zituela pentsatzen baitzen). Gag proteinatik proteolitikoki honako hauek eratortzen dira: MA proteina (matrizea), CA proteina (kapsida), NC proteina (nukleokapsida) eta, batzuetan, funtzio ezezaguna duten beste proteina bat (Vogt, 1997).

Oro har, onartzen da MA proteina mintz-proteina periferikoa dela eta uste da mintzaren eta MA proteinaren arteko elkarrekintza egon daitekeela. Proteina horren funtzioa guztiz argituta ez badago ere, miristilatzen dela ikusi da eta, prozesu hori erretrobirusa muntatzeko beharrezkoa da. Halaber, ikusi da, birus ezberdinetako MA proteinak lipido ezberdinekiko afinitate desberdina dutela. Bestalde, badirudi, *env* genearekin elkarrekintzaren bat dutela (Vogt, 1997).

CA proteina antigenikoa da. Proteina horren funtzioa guztiz argitu ez bada ere, badirudi birusaren RNA gordetzen duen konplexu erribonukleikoaren inguruan “maskor” bat eratzen duela, kapsida alegia. Hala ere, “maskor” horren funtzio biologikoa ez dago argi. Ikerketa batzuek aditzera eman dute CA proteinak infekzioaren hastapenean garrantzia izan dezakeela (Vogt, 1997).

Bestalde, ikusi da NC proteinak RNA genomikoarekiko lotura estua duela eta badirudi proteina horrek erretrobirusaren muntaketan parte har dezakeela. Orobat, ikusi da tRNA PBS guneari lotzea erraztu eta alderantzizko transkripzioa suspertzen duela (Vogt, 1997).

1.1.2. *pro-pol* geneak

pro geneak PR (proteasa birikoa) proteina kodetzen du. Proteina horrek *gag*, *pro*, *pol* eta, batzuetan, *env* geneek kodetutako proteinak proteolitikoki prozesatzen ditu eta, prozesu horren ondorioz, proteina horiek funtzionalak bilakatzen dira. Horrela, birusaren muntatzearen bukaeran, proteasak aldaketa morfologikoak eragiten ditu birusean eta infekzioso bihurtzen du (Vogt, 1997).

pol geneak RT (alderantzizko transkriptasa) eta IN (integrasa) entzimak kodetzen ditu. RT entzimak DNA polimerasa jarduera du; RNA molde bat erabiliz, DNA sekuentziak sintetizatzen ditu. Halaber, RT entzimak RNAsa H jarduera du; hots, RNA/DNA hibridoak degradatzen ditu (Vogt, 1997).

IN delako entzimak genoma erretrobiralaren erreplikazioan parte hartzen duelarik, erretrobirusen DNA genomaren txertatzeko proteina gakoa da. Proteina hori gai da sintetizatu berri diren erretrobirusaren DNA harizpiak jasotzeko eta genomaren DNA ituan zoriz txertatzeko (Vogt, 1997).

1.1.3. *env* genea

env geneak zelulako proteina hartzaileekin espezifikoki elkarrekintza duten SU (gainazal glikoproteina) eta TM (transmintza proteina) proteinak kodetzen ditu. Horrelako proteinak estalkia eratzen dute, infekta daitezkeen zeluletan birusa adsorbatzen eta sartzen laguntzen duen egitura (Vogt, 1997).

1.1.4. *Gainontzeko geneak*

Erretrobirus gutxi batzuek kokapen genomiko aldakorra duen *dat* delako genea azaltzen dute, eta horrek DU proteina (deoxiuridina trifosfata) kodetzen du. DU proteinak deoxiuridina trifosfata birioian sartzea ekiditen duelarik, kieszentzian dauden zeluletako erreplikazioa errazten du. Gainontzeko geneen artean birusaren adierazpena erregulatu eta koordinatzen dituzten geneak ere aurki daitezke (Vogt, 1997).

Erretrobirus batzuek *onc* geneak (onkogeneak) eduki ditzakete beren egitura genomikoan. Horrelako geneak daramatzaten erretrobirusek tumoreak sortzen dituzte eta, oro har, ez dira funtzionalak, erreplikaziorako beharrezkoak diren geneetan ezabatzeak pairatu baitituzte *onc* geneak kokatzeko. Erreplikazioa gertatu ahal izateko, *onc* genea duten birus horiek beharrezkoa dute funtzionala den erretrobirus baten laguntza (Vogt, 1997).

1.1.5.LTRak

LTRak erretrobirusen alboetan dauden egiturak dira. LTRak sekuentzia berdin-berdinak dira eta tamaina aldakorra duten eskualdeez osatuta daude: U3 (*unique 3'*), R (*repeat*) eta U5 (*unique 5'*) eskualdeak. Guztira LTRek 300 eta 1800 base-pareko luzera dute. Transkripzioaren kontrol-elementu gehienak U3 eskualdean kokatuta daude, adibidez, sustatzaileak eta areagotzaileak. Areagotzaileak berebiziko garrantzia du infektatuko den ehuna zer-nolako izango den, hau da, patogenesia nolako izango den baldintzatzen baitu. PoliA isatsa seinalea R eta U5 eskualdeen artean kokatuta dago (Vogt, 1997).

LTRak genoman txertatzeko orduan, berdin-berdinak dira eta haien dibergentzia txertatzen denbora estimatzeko erabiltzen da (Dangel et al., 1995; Vogt, 1997; Hughes & Coffin, 2005).

1.2. ERVen sailkapena eta banaketa

Erretrobirus endogenoen (ERV) eta erretrobirus exogenoen (XRV) sailkapena ez da batera garatu, beren artean erlazioa egonda ere. Horren ondorioz, oro har, ERVak ez dira XRVak sailkatzen diren generoetan barneratzen. Gainera, hasieran gizakien (*Homo sapiens*) ERVak (HERV, *human endogenous retrovirus*) sailkatzeko erabilitako nomenklaturak lan hori oztopatu du. Gizakiaren ERVen taldeak familia moduan izendatzen ziren beren PBSaren (hasle-loturaren gunea) arabera. HERVak izendatzeko familia terminoa erabiltzeak nahasmena sortzen du, batetik, erretrobirusek Retroviridae familia osatzen dutelako, eta, bestetik, ikusi delako PBS berdina duten erretrobirus batzuek ahaidetasun filogenetikorik ez dutela. Hala ere, ERV taldeak izendatzeko familia terminoa erabili ohi da (Tristem, 2000; Griffiths, 2001).

Azken lanetan, gizakirako ezarri den irizpide berri bati jarraituz (Gifford et al., 2005), ERVak hiru taldetan sailkatzen dira XRVekin duten antzekotasunaren arabera: I klasea, XRVen *Gammaretrovirus* eta *Epsilonretrovirus* generoen parekoak; II klasea, *Alpharetrovirus*, *Betaretrovirus*, *Deltaretrovirus* eta *Lentivirus* generoen parekoak; eta III klasea, *Spumavirus* generoarekin parekoak (Gifford et al., 2005). Familien eta klaseen izendapenaz gain, Blombergekin et beste zenbaitek egindako lanean (2009), ERV indibidualak sailkatzeko eta izendatzeko gomendioak egin dira. Hor proposatzen dute ERVak izendatzerakoan ERVen antzekotasuna, ezaugarri estrukturalak, funtzioak eta antzinako izendapenak kontutan hartu behar direla (Blomberg et al., 2009).

ERVen banaketa aztertzeke, Robert Gifford-ek eta Michael Tristem-ek 2003an argitaratutako berrikuspena erabiliko dugu oinarri gisa (Gifford & Tristem, 2003).

ERVen bilaketak ornodunengan nahiz ez-ornodunengan egin badira ere, ERVak ornodunengan bakarrik aurkitu dira (I-3. irudia). Oinarritzko ornodunengan izan ezik, Cephalaspidomorphi (lanproiak) eta Myxini klaseetan alegia, aztertutako ornodun klase guztietan agertzen dira ERVak. Guztira, ERVak sei klaseren barnean dauden 25 ordenetan agertzen dira, kartilagodun arrainetatik mamalioetara. Hala ere, mamalioak gehien ikertu diren espezieak dira eta, beraz, ERVen banaketari buruz dakiguna alboratuta egon daiteke. ERV antzeko elementuak *Drosophila* eta landareetan aurkitu dira, baina analisi filogenetikoek adierazten dute erretrotransposoietatik gertuago daudela (Herniou et al., 1998; Gifford & Tristem, 2003).

Letra lodiz XRVak; gainontzekoak, ERVak. (Herniou et al., 1998)-tik moldatua.

1.2.1.1 Klaseko ERVak

I klaseko ERVak ostalari-hein zabalena duten ERVak dira (Herniou et al., 1998).

Gammaretrovirus generoko XRVeK egitura sinplea dute eta, erretrobirus generoen artean, ostalari-banaketaren hein zabalena (Gifford & Tristem, 2003). Horiekin erlazionatutako ERVak anfibioetan, narrastietan, hegaztietan eta mamalioetan aurkitu dira. Arrainetan eta beste oinarrizko kordatueta, aldiz, ez dira aurkitu, nahiz eta bilaketak egin (Martin et al., 1999).

Epsilonretrovirus generoko XRVek egitura genomiko konplexua dute eta arrainetan soilik aurkitu dira. Salbuespenak salbuespen, genero horretako XRVkin erlazionatutako ERVrik ez da topatu. Hala

ere, *Xenopus laevis* apo espeziean, XRV horiekin erlazio urruna duen ERV bat aurkitu da (Gifford & Tristem, 2003).

1.2.2. II Klaseko ERVak

II klaseko ERVak hein murriztua azaltzen dute: hegaztietan eta mamalioetan bakarrik agertu dira (Gifford et al., 2005).

Alpharetrovirus generoko XRVak hegaztien *Gallus* generoan aurkitu dira batez ere. Egitura genomikoa sinplea badute ere, batzuetan, genomaren zatiak galtzen dituzte onkogeneak eramateko. Salbuespen bakarra RSV (*Rous sarcoma virus*) birusa da, non onkogeneek ez duten ohiko generik ordezkatzeko. Genero horrekin erlazionaturiko ERVak, hasieran, oiloan (*Gallus gallus*) bakarrik aurkitu ziren. ALVE familia ERV horietako talde bat da. Familia horren txertatzeak oiloaren aitzindari basatian aurkitu dira, baina ez beste Galliformes-etan, eta beraz, txertatzea orain dela gutxi gertatu dela esan nahi du. EAV-0 ERV familia ere *Gallus* generoan agertzen da barreiatuta. Familia horrek *Gallus* generoaren filogeniari estuki jarraitzen dio eta, hortaz, txertatzea antzinakoa dela ondorioztatu da. Hala ere, badaude hainbat ikerketa ERV banaketa *Gallus* generoan zabalagoa izan daitekeela aditzera ematen dutenak; adibidez, *Bonasa umbellus* lagopodo espeziean, TERV izendatutako ERVa aurkitu daiteke. Geroago egindako ikerketetan, *Alpharetrovirus* generoarekin erlazioa duten ERVak 15 hegazti ordenatan ere detektatu ziren (Gifford & Tristem, 2003; Gifford et al., 2005).

Betaretrovirus generoko XRVak mamalioetan besterik ez dira aurkitu (karraskarietan, hausnarkarietan, primateetan, eta abarretan) eta gehienek genoma sinplea dute. Genero horretako hainbat erretrovirusek forma exogenoa eta endogenoa dute, adibidez, *Jaagsiekte sheep retrovirus* (JSRV) ardietan (*Ovis aries*) eta *Mouse mammary tumor virus* (MMTV) saguetan (*Mus musculus*). Era berean, forma exogenorik ez duten ERVak agertu dira. Genero horrekin ahaidetutako ERV horiek mamalioetara (primateetara, karraskarietara, eta abarretara) mugatuta daude. ERVek, mamalioetan, banaketa zabala eta, era berean, irregularra dute, aztertu ziren mamalioen 49 taxoietatik 25etan detektatu baitziren. Karraskarietan, ERV horietaz gain, ERVekin modu ahulean filogenetikoki multzokatzen diren IAP elementuak (*intracisternal A-type particle*) agertu dira (Herniou et al., 1998; Gifford & Tristem, 2003; Gifford et al., 2005).

Deltaretrovirus generoko XRVak primateetan eta behietan bakarrik aurkitu dira, egitura konplexua dute eta, batzuen egituran, onkogeneak agertzen dira. Ez da horiekin lotutako ERVrik aurkitu (Gifford & Tristem, 2003).

Lentivirus generoko XRVak primateetan, felidoetan eta hainbat ungulatutan (ahuntzan, ardian, behian eta zaldian) aurkitu dira. Genoma konplexua dute eta batzuek onkogeneak daramatzate (Gifford & Tristem, 2003). *Deltaretrovirus* setan bezala, erlazionaturiko ERVrik ez zegoela pentsatzen zen, baina ikerketa berriek *Lentivirus* en antzeko ERVak aurkitu dituzte untxian (*Oryctolagus cuniculus*) (Katzourakis et al., 2007b) eta *Microcebus murinus* lemure espeziean (Gifford et al., 2008).

1.2.3. III Klaseko ERVak

Spumavirus generoko XRVek genoma konplexua dute eta mamalioen (felidoetan, behietan, primateetan eta abarretan) artean, oso zabalduta daude. Ez da horiekin erlazio esturik duen ERVrik aurkitu. Hala ere, erlazio ahul bat azaltzen duten ERVak aurkitu dira mamalioetan, hegaztietan, narrastietan eta anfibioetan. Egiturak konparatuz, ERV urrun horiek *spumavirusek* dituzten bi gene gehigarri galdu dituzte (Benit et al., 1999; Gifford & Tristem, 2003).

Mamalioen artean, *Spumavirus* generoarekin erlazionatutako ERVek banaketa ezberdina dute. Southern-blot teknikaren bidez egindako ikerketan (Benit et al., 1999), tximinoetan eta saguetan seinale oso indartsua ikusi zen; aztertutako gainontzeko mamalioetan (katua, txakurra, untxia, behia, txerria, zaldia eta arratoia), berriz, seinaleak ahulak izan ziren. Ez zen seinalerik detektatu hegaztietan, arrainetan eta ez-plazentadun mamalioetan. Aztertutako mamalio guztietan agertu denez, *Spumavirusekin* erlazionatutako ERVak antzinako ERV batetik eratorri zirela ondorioztatu da.

Bestalde, tximinoetan eta saguetan ikusten den seinale indartsua ERVen kopuru emendioarekin erlazionatu da; primateen kasuan mundu zaharreko eta berriko tximinoak banatu zirenean gertatu zen eta, saguen kasuan, duela gutxi gertatu zen. ERV horien erlazio filogenetikoari dagokionez, tximinoen ERVek talde bat osatzen dute; saguenek beste bat eta, gainontzeko mamalioen ERVek ez dute taldekatze sendorik azaltzen (Benit et al., 1999).

1.2.4. Gainontzeko ERVak

ERV horietaz gain, zalantzazko kasu batzuk badaude. XRVekin erlazorik gabeko ERV elementuak krokodiloetan (*Crocodylus* sp.) eta pitoi sugeetan (*Python molurus*) detektatu dira (Gifford et al., 2005). ERV horiek gainontzeko ERVekin antzekotasun-maila txikia badute ere analisi filogenetikoek ERVen lerro filogenetikoaren partaideak direla ondorioztatzen dute. Pitoi sugearen ERVaren kasua harrigarria da; II klasekoekin multzokatzen da, baina mamalio- edo hegazti-jatorria ez duen ERV bakarra da. Gainera, ERV hori ez da beste boa-sugeetan detektatu (Gifford & Tristem, 2003; Gifford et al., 2005).

1.3. ERVen *in silico* azterketa

Genomikaren, hots, genomen sekuentziazioaren aroan, erretrobirus endogenoekin (ERV) lan egiteko informazio ugari dugu eskuragarri. Genoma osoen sekuentziak ditugunez, errezagoa da genoma jakin bateko ERV sekuentzien bilaketa, banaketa, dibertsitatea eta dinamika ikertzea, eta, horri esker, hainbat genomatan gertatzen dena konpara daiteke (Gifford & Tristem, 2003). ERVen detekzioa eta karakterizazioa egiteko lanabes bioinformatikoak garrantzitsu bilakatu dira. Orain arte, ERVen bilaketa bioinformatiko sakonak primateen, karraskarien eta oiloen genomatan soilik egin dira; geroago ikusiko ditugu horiek.

1.3.1. Bilaketa programak

Azken urteetan ERVak detektatzeko hainbat estrategia erabili dira, eta lan horretarako espresuki diseinatutako programak ere sortu.

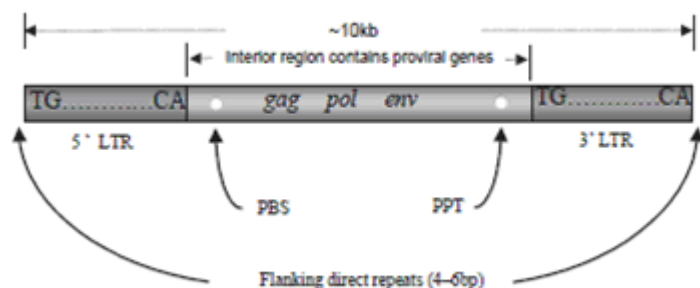
Genomak sekuentziatzen direnean, elementu errepikakorrak detektatzeko, hala nola ERVak, RepeatMasker programa erabiltzen da Genoma Proiektu gehienetan. Programa horrek genomatan elementu horien antzeko sekuentziak bilatzeko Repbase datu-basea erabiltzen du. Repbase genoma eukariotikoetan dauden elementu errepikakorren sekuentzia prototipikoak gordetzen dituen datu-basea da. ERVei dagokienez, Repbasek hainbat espezieetan agertzen diren 2000 ERV *loc*itik gora biltzen ditu (Jurka, 2000; Smit et al., 2004; Blikstad et al., 2008).

ERVen antzeko sekuentziak bilatzeko BLAST algoritmoa askotan erabili da. BLAST algoritmoak bilaketa heuristikoa erabiltzen ditu sekuentzia jakin bat sekuentzia askoren artean (sekuentzia datu-base bat, adibidez) aurkitzeko. Bilaketa horiek modu lokalean egiten ditu; hau da, sekuentzia osoa bilatu beharrean, sekuentziaren zati bat bilatzen du, antzekotasun gehiago topatzeko. Lehenengo amu-sekuentziaren azpisekuentzia zerrenda bat egiten du eta horiek bilatzen ditu. Geroago, aurkitutako azpisekuentziak erabilitako parametroen arabera, ahal duen guztia luzatzen ditu. Azkenik, azpisekuentzia horietan oinarrituz, lerrokatze onena muntatzen du eta lerrokatzeari buruzko kalkuluak egiten ditu. Ezagutzen diren erretrobirusen sekuentziak erabiliz (bai endogenoak, bai exogenoak), bilaketak egin dira genomatan, batez ere *pol* geneko RT eskualdea erabiliz (Altschul et al., 1997; Madden, 2002; Gifford & Tristem, 2003; Sperber et al., 2007).

Azken urteetan LTRdun elementuak eta ERVak detektatzeko hainbat programa sortu dira. Hemen, LTR_STRUC (McCarthy & McDonald, 2003) eta RetroTector© (Sperber et al., 2007) programak aztertuko ditugu, mamalioen zenbait genomatan ERVak detektatzeko erabili direlako. Hala ere, Genome Parsing Suite (McClure et al., 2005) programa sortu da, besteak beste, “agente erretroideak” detektatzeko, eta LTRak bilatzeko egindako LTR_par (Kalyanaraman & Aluru, 2005) edota LTRharvest (Ellinghaus et al., 2008) programak.

I.3.1.1. LTR_STRUC

LTR_STRUC LTRdun egiturak espresuki bilatzen dituen programa da. Sekuentzia genomikoak aztertzean programa horrek distantzia hein jakin baten barnean dauden balizko LTRak bilatzen ditu (I-4. irudia). Balizko LTR horiek detektatu ondoren, aztertzen du ea LTR horien artean PBS, PPT eta LTRen alboetan ohikoak diren nukleotidoak badauden (McCarthy & McDonald, 2003).

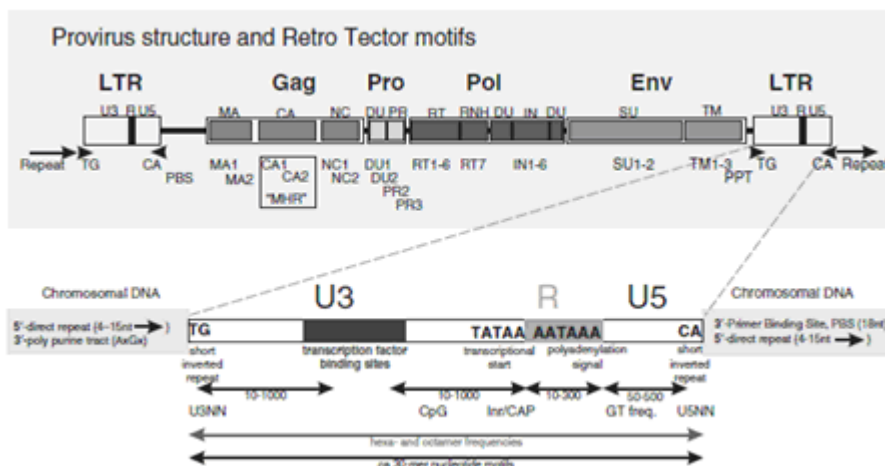


I-4. irudia: LTRdun elementuak detektatzeko LTR_STRUC-ek bilatzen dituen egiturak (McCarthy & McDonald, 2003).

sekuentzia luzatu horietan, LTRen bukaera-puntuak zehazten ditu. LTRak definitu ondoren, haien arteko eskualdean lehen aipatutako ezaugarri erretrobiralak aztertzen ditu (McCarthy & McDonald, 2003).

I.3.1.2. RetroTector©

RetroTector© programa ERVak espezifikoki detektatzeko sortutako programa da. Programa horrek, hainbat pausotan zehar, ERVen bilaketa sakona egiten du. Lehenengo, genomak, balizko



I-5. irudia: ERVen bilaketa egiteko RetroTector©-ek erabiltzen dituen motiboak (Sperber et al., 2007).

detekzioa kontserbatutako motiboen, motiboen arteko distantzia-mugen eta aminoazidoen antzekotasun- matrizeen mendekoa da. Motibo bakoitza genero erretrobiral batekin lotzen du eta motiboen arteko distantziak erretrobirus ezagunetan oinarritzen dira; hala, segurtasun-tarte bat gehitzen dio. Distantzia horien barruan dauden motiboak modurik sinesgarrienean konbinatzen dira balizko ERVa prozedura heuristikoki baten bidez definituz. Hala ere, distantzia horietatik kanpo gelditzen diren motiboak onar ditzake, txertatzeak eta ezabatzeak kontutan hartzeko. Balizko ERV hori genero bati egokitzen zaio eta puntuazio bat eman, generoaren, luzeraren eta motiboen egokitzapenaren arabera (Sperber et al., 2007).

Detekzioa aurrera eramateko hainba irizpide erabiltzen dira: 1) Bilaketa hierarkikoa da: erabilitako motiboak azpigenetean multzokatzen dira (I-5. irudian ikus daitezkeenak), bakoitzean motiboen

RetroTector©-en

bilaketa eta balizko azpigeneak definituz; 2) azpigene sekuentzian, hutsune-luzera ezberdina ezartzen da; 3) motibo batzuk bilaketak fintzeko soilik erabiltzen dira; 4) sekuentzia luzeak zati txikiagotan banatzen dira prozesatu baino lehenago, baina ERVak galtzeko arriskurik gabe (Sperber et al., 2007).

Sekuentzien estatistikoak kalkulatzeko, ERVen bi ezaugarri erabiltzen ditu. Alde batetik, gene kodetzaileen (*gag*, *pro*, *pol* eta *env*) sekuentzian, hexamero jakin batzuek (sei nukleotidoko multzoak) ORF irakurtarauan beste irakurtarauetan baino proportzio handiagotan azaltzen dituztela. Bestetik, LTRetan, oktamero jakin batzuk ohikoagoak direla (Sperber et al., 2007).

Azkenik, lerrokatzeak programazio dinamikoaren bidez egiten ditu, prozedura estandarrek erabiliz (Sperber et al., 2007).

Bestalde, RetroTector©-en sortzaileek RetroTector© online lanabesa eskuragarri jarri dute, 5 eta 10000 Kbp arteko sekuentziak analizatzeko gai den web bidezko lanabesa da (Sperber et al., 2009).

1.3.2. Genoma osoko ERVen bilaketak

Horago aipatu bezala, bilaketa programa horiek, primateen, karraskarien eta oiloen genomak aztertzeke erabili dira ERVen karakterizazio sakona egiteko asmoz (I-1. taula).

I-1. taula: ERVen *in silico* bilaketak egindako genomak.

Atal honetan aipatu diren adibideetan oinarrituta.

Espezia	Metodoa	Genomaren %	Familiak	Erreferentzia
Gizakia	Repeatmasker	%4,64		(Internation Human Genome Sequencing Consortium, 2001)
	BLAST (genomaren %7)	22 ERV talde	22	(Tristem, 2000)
	RepeatMasker	%2,7	39	(Paces et al., 2002)
	BLAST	%1,1 (7.836 ERV)		(Villesen et al., 2004)
Txinpantzea	LTR_STRUC+BLAST		42	(Polavarapu et al., 2006)
Makakoa	RepeatMasker		8 espezifiko	(Han et al., 2007)
Sagua	RepeatMasker	%4,4		(Mouse Genome Sequencing Consortium, 2002)
	BLAST		II. klaseko 7	(Baillie et al., 2004)
	LTR_STRUC		21	(McCarthy & McDonald, 2004)
Arratoia	RepeatMasker	%5,05		(Rat Genome Sequencing Project Cons, 2004)
	BLAST		II. klaseko 7	(Baillie et al., 2004)
Oiloa	LTR_STRUC+BLAST	%2,9	14	(Huda et al., 2008)

1.3.2.1. Gizakia

Genoma osoko ERV bilaketa sakonena gizakiaren genomaren egin da. Gizakiaren ERVak bilatzeko eta aurretik zegoen informazio esperimentalarekin uztartzeko estrategia ezberdinak erabili dira. Gizakiaren genoma sekuentziazterakoan, Repeatmasker-ekin egindako azterketan, ERVek gizakiaren %4,64ko estaldura zutela ondorioztatu zen (I. klasekoek %2,89, II. klasekoek %0,31 eta III. klasekoek %1,44) (Internation Human Genome Sequencing Consortium, 2001); hala ere, kopuru hori %8ra hel daitekeela uste da (Gifford & Tristem, 2003).

Lehenengo genoma osoko hurbilketa gizakiaren genomaren %7arekin egin zen, garai hartan eskuragarri zegoen informazioa erabiliz (Tristem, 2000). Ezagunak ziren ERVen *pol* geneko RT eskualdea erabiliz BLAST bilaketak egin ziren gizakiaren genomaren estaldura horretan. Analisi filogenetikoek ERVak 22 familiatan sailkatu zituzten, gehienak I klasekoak, eta lehendik ezagutzen ez ziren beste 6 familia berri detektatu ziren.

Gizakiaren genoma osoarekin egindako lehen azterketan, RepeatMasker erabili zen (Paces et al., 2002). Programa horrekin elementu errepikakor guztiak detektatzen direnez, bigarren pausoa gizakiaren ERVak iragazi eta sailkatu ziren, eta HERVd datu-basea sortu zen (<http://herv.img.cas.cz/>). Guztira, 39 ERV familia deskribatuta daude datu-base horretan, eta gizakiaren genomaren %2,7 ordezkatzen dute.

Bestalde, gizakiaren genomaren ERVen gene nagusiak (*gag*, *pro-pol* eta *env*) BLAST bidez bilatu ziren (Villesen et al., 2004). Hurbilketa horren bidez, gizakiaren genomaren 7836 ERV identifikatu ziren. ERV horiek gizakiaren genomaren %1,1 ordezkatzen dute, eta ikusi zuten banaketa kromosomiko homogeneorik ez zutela. ERV hauek eta detektatu ziren ia 38000 geneak Retrosearch datu-basea elikatzeke erabili ziren (<http://www.retrosearch.dk/>).

I.3.2.2. Gainontzeko primateak

Txinpantzearen (*Pan troglodytes*) genomak ERVak detektatzeko, lehenengo, LTRdun egiturak bilatu ziren LTR_STRUC programa erabiliz eta, programa horren emaitzetan oinarrituta, *pol* genearen RT eskualdearen antzeko eskualdeak BLAST programa erabiliz bilatu ziren txinpantzearen genomak. Detektatutako ERVak 42 familiatan sailkatu ziren analisi filogenetikoen arabera eta familia gehienak I. klasean kokatu ziren. Detektatutako ERVak 7.000 eta 10.000 base-pare arteko tamainakoak ziren. Txertatze-denborari dagokionez, I. klasekoak orain dela 0,8-82,9 milioi urte artean txertatu ziren; II. klasekoak, 2-97 milioi urte artean; eta III. klasekoak, 30-145 milioi urte artean. CERV1/PTERV1 eta CERV2 delako taldeak izan ezik, beste talde guztiek talde homologo bat dute gizakiarengan (Polavarapu et al., 2006).

Makakoaren (*Macaca mulatta*) kasuan, Mundu Zaharreko Tximinoen espezifikokoak diren ERVak aztertu ziren. Gizakiarengan eta txinpantzearen eskualde homologoetan agertzen ez ziren eta makakoan agertzen ziren elementu errepikakorrak bilatu zituzten. Horretarako, 3 bideko multi-Z lerrokatzea egin zen BLAT programa erabiliz, eta elementu errepikakorrak RepeatMasker erabiliz detektatu ziren. ERVei dagokionez, gehienak hominido eta mundu zaharreko tximinoak banatu baino lehenagotik finkatuta zeuden genomak. Banaketaren ostean, makakoaren genomak, 8 familia txertatu dira, 5 I. klasekoak eta 3 II. klasekoak. Oro har, ERV gehienak I klasekoak dira (Han et al., 2007).

I.3.2.3. Karraskariak

Saguaren genoma sekuentziatu zenean, Repeatmasker-ekin egindako azterketaren bidez ikusi zen, ERVek saguaren %4,4a ordezkatzeko dutela (I. klasekoak %0,68, II. klasekoak %3,14 eta III. klasekoak %0,58) (Mouse Genome Sequencing Consortium, 2002). Bestalde, arratoiaren kasuan, ERVek genomaren %5,05 ordezkatzeko dute (I. klasekoak %0,97, II. klasekoak %3,24 eta III. klasekoak %0,84) (Rat Genome Sequencing Project Cons, 2004). Karraskarien genomekin bi azterketa nagusi egin dituzte: bata saguaren eta arratoiaren genomak, II. klaseko ERVak bilatuz, eta bestea saguaren genomak, LTR-erretrotransposoiak bilatuz.

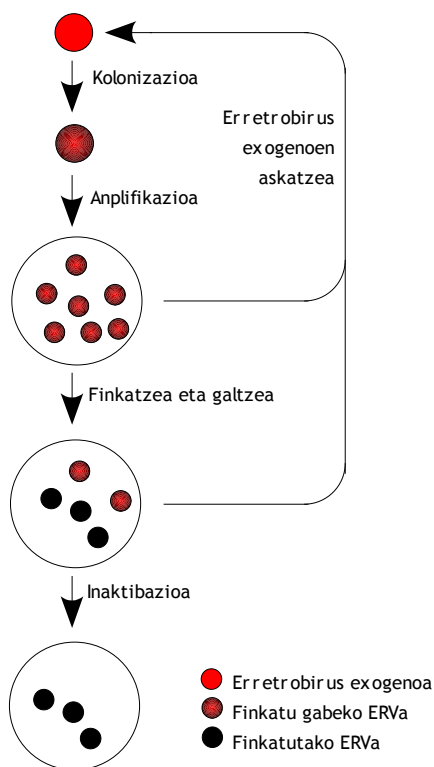
II. klaseko ERVak karraskarien genomak dauken ERVen gehiengoa dira. ERV mota horiek saguaren eta arratoiaren genomak aztertzeko, sekuentzia-antzekotasunean oinarritu ziren, hots, BLAST erabili zuten. Metodologia horren bidez, *pol* genearen RT eskualdea karraskari horien genomak bilatu zen. Detektatu ziren ERVen analisi filogenetikoek II. klaseko 7 familia zeudela adierazi zuten; genomaren jatorria berdina zion, bi espezie horietan agertzen diren familiek jatorri bera baitzuten (Baillie et al., 2004).

LTR-erretrotransposoiei dagokionez, LTRdun egiturak bilatzen dituen LTR_STRUC programa erabili zen detekzioa egiteko. Detekzio hori osatzeko, RT eskualdeen BLAST bidezko bilaketak egiten ziren. Analisi filogenetikoek ondorioztatzen dute ERVen 3 klaseak ordezkatzeko dituzten 21 familia daudela eta ERV gehienak II. klasekoak direla. Oro har, detektatutako LTR-erretrotransposoiak 2.700 eta 7.200 bp bitarteko tamaina dute (McCarthy & McDonald, 2004).

I.3.2.4. Oiloa

ERVen genoma osoko azterketetan, oiloarena izan da aztertutako lehen hegazti-genoma (Huda et al., 2008). ERVak detektatzeko, txinpantzearekin erabilitako prozedurari jarraitu zitzaion: lehenengo, LTRdun elementuak bilatu ziren LTR_STRUC programa erabilita. Bigarren pausuan, bilaketa horretatik lortutako *pol* genearen RT eskualdeak oiloaren genomak BLAST bilaketak egiteko amusekuentziak bezala erabili ziren. Detektatutako ERVek oiloaren genomaren %2,9a ordezkatzeko dute. Hala detektatutako ERVen analisi filogenetikotik ondorioztatu zen guztira 14 ERV familia zeudela, zeinetatik 11 talde ezezagunak ziren. Taldeen txertaketari dagokionez, zenbatetsi zen familiarik gazteena orain dela 3-0 milioi urte artean txertatu zela eta zaharrena orain dela 136,5-143,9 milioi urte (Huda et al., 2008).

I.4. ERVen dinamika ebolutiboa



I-6. irudia: ERVen bizi-zikloa.
(Gifford & Tristem, 2003)-tik moldatua.

Erretrovirus endogenoak (ERV) genoma ostalariaren finkantzea hainbat faktorek baldintzatzen dute. Besteak beste, honako hauek aurki ditzakegu: zelula guztiek dituzten mugak gainditzea, hozi-lerroa heltzeko ugaltze-ehunak infektatzea, infektatutako hozi-lerroaren biziraupena baimentzea, ostalariaren ondorengoak egokitzapena galdu gabe bizirautea eta finkatu eta gero ostalariaren populazioaren itoguneak gainditzea (Blikstad et al., 2008). ERVak ostalariaren txertatu eta gero, genomaren parte direnez, genomaren edozein sekuentziak jasaten dituen presio ebolutiboak jasaten dituzte, hala nola, mutazioa, hautespena, errekonbinazioa, eta abar. Hainbat ERVek aurkako hautespena jasan dezakete eta populazio ostalaritik desagertu. Bestalde, genomaren txertatzeak ostalariaren eragin negatiboak sor ditzake, ostalaria kaltetu eta, ondorioz, ERVa ez da hurrengo belaunaldietara transmitituko. Hala ere ERVek populazio ostalariaren genomak kolonizatzen dituzte, populazio ostalariaren jito genetiko eta arraste genetiko direla-eta, alde batetik, eta ERVek erreplikatzeko gaitasuna mantentzen dezaketelako beren kopurua genoma ostalariaren emendatuz, bestetik. Prozesu horien ondorioz, finkatutako eta ez-finkatutako elementuez osatutako ERV familia monofiletiko bat sortzen da (ikus I-6. irudia). ERV familia gehienak populazio ostalariaren genomatik desagertzeko joera dute, bai ostalariaren izan dezaketen eragin patogenikoengatik, bai ERVek jasan

dezaketen mutazioengatik. Hala ere, hainbat ERV familia ez dira desagertzen, alde batetik familia horiek funtzionaltasuna mantentzen dutelako eta, bestetik, familia horien adierazpen kontrolatuak indibiduoetako abantailaren bat eman diezaikeelako (Gifford & Tristem, 2003; Jern & Coffin, 2008; Blikstad et al., 2008).

I.4.1. Genomen kolonizazioa

ERV familia bat sortzeko lehen pausoa XRV bat hozi-lerroko zelula bateko genoman txertatzea da, zelula ostalariaren defentsak gainditu eta gero.

I.4.1.1. Txertaketa

Erretrovirusaren txertaketa erretrovirusaren gainazaleko glikoproteinak zelularen hartzaileei lotzen zaizkienean hasten da. Lotuz gero, erretrovirusa zelularen mintzarekin fusioan sartzen da eta birioiaren muina zitoplasman askatzen da. RNA erretrobiralak, alderantzizko transkripzioaren bidez, harizpi bikoitzeko DNA bilakatzen da. DNA erretrobiral hori birioiaren muintetik eratorritako proteinez osatutako egitura batekin batzen da eta txertaketa aurreko konplexua (PIC, *preintegration complex*) eratzen dute. Ondoren, DNA erretrobiralak nukleora garraiatzen da eta ostalariaren genoman txertatzen da birusak kodetzen duen integrasa proteinaren bidez (Jern & Coffin, 2008).

Erretrovirusak genomatan txertatzea ez da guztiz zorizko gertaera. Hainbat XRVekin egindako ikerketetan, genomatan txertatzerakoan, XRV bakoitzak tropismo ezberdina azaltzen du: GIBa (giza immunoeskasiaren birusa, gizakien birusa), batez ere, gene aktiboetan txertatzen da eta MLVa (*murine leukemia virus*, saguen birusa) geneen 5' bukaeran; ASLVetan (*avian sarcoma leukosis virus*, oiloen birusa), berriz, ez da lehentasunik hauteman (Bushman et al., 2005).

I.4.1.2. Ostalariaren defentsak

Ostalariaren genoman ERVak finkatzea oztopatzen duten faktoreen artean (I-7. irudia), elementu

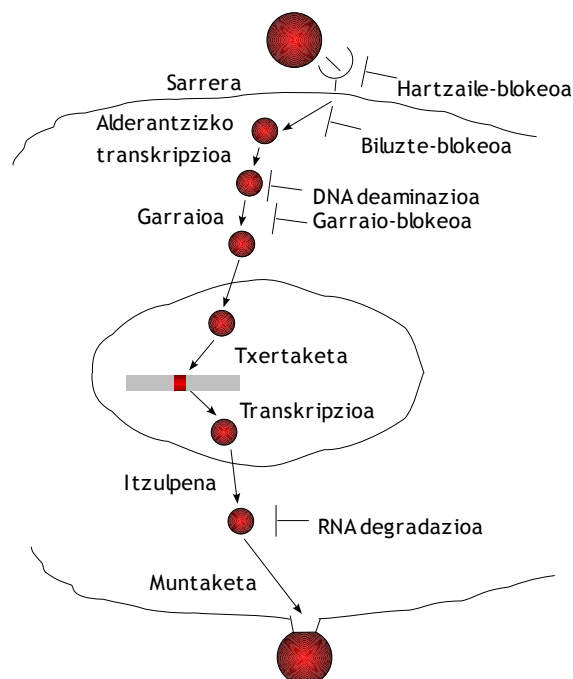
transposakorren eta azido nukleiko kanpotarren aurkako defentsak aurki ditzakegu, eta faktore horiek dute eragin handiena ERVen eboluzioan. Besteak beste, zelularen hartzaileen interferentzia, beste ERV batzuen interferentzia (5.2 atalean aztertuko dugu) edota murrizte-faktore erretrobiralak aurki ditzakegu (Blikstad et al., 2008).

Erretrobirus bat ez sartzeko ezaugarririk garrantzitsuen erretrobirusen hartzaileak egotea da. Horrelako hartzaileak egotekotan, hartzailea blokeatu daiteke erretrobirusa ez sartzeko lehen defentsahesi modura. Saguarengan, funtzio hori duen Fv4 genea deskribatu da. Gene horrek MuLV birusa sartzeko ekiten du, erretrobirus horren hartzailea blokeatuz. Gainera, uste da Fv4 genea, berez, *gag* eta *pol* geneak galdu dituen ERV baten *env* gene akasdu batetik eratorri dela (Goff, 2004).

Erretrobirusak sintetizatutako DNA berriari erasotzea beste defentsahesi bat da. Hesi horretan zitidina deaminatzen duen APOBEC3G genea kokatzen da. Halaber, ikusi zen gene horrek gizakian *vif* gene gabeko GIB birusetan DNA (-) harizpian zitosenen %4a urazilora aldatzen dituela eta, horren ondorioz, (+) harizpian guaninak adenina bihurtzen direla. Aldaketa horien ondorioz, DNA erretrobiralak ez-egonkorra bilakatzen da eta hainbat entzima bidez degradatu daiteke, edo sekuentzia hain mutatu gelditzen da, ezen birusak funtzionaltasuna galtzen duen. Gene hori beste tximino batzuetan ere aurkitu da antzeko ezaugarriekin. APOBEC proteina-familiako APOBEC3F eta APOBEC3B proteinek ere antzeko funtzioa dute (Goff, 2004).

DNA erretrobiralaren sintesia eta zelula barruko zirkulazioa blokeatzea hurrengo defentsahesizat har ditzakegu. Hor Fv1 eta TRIM5 α geneak koka daitezke. Saguetan, Fv1 geneak (ERV baten *gag* genetik eratortzen da), PICari (integrazio aurreko konplexua) erasotzen dio edo harrapatu egiten du, eta ez dio nukleoan sartzen uzten. Gizakiengan, Ref1 geneak Fv1 genearen antzeko funtzioa du. Haien arteko ezberdintasun bakarra honako hau da: Ref1 genearen eragina alderantzizko transkripzioa baino lehenago gertatzen dela (Goff, 2004). Bi gene horien funtzioa bermatzeko, TRIM5 α genea beharrezkoa da. Gene hori erretrobirusen ostalari-heinaren faktorerik garrantzitsuen dela aditzera eman da eta, espezieen arteko transmisioan, hesi garrantzitsutzat har daiteke (Goff, 2004).

Azken hesi gisa, erretrobirusaren adierazpena blokeatzen dituzten mekanismoak aurki ditzakegu. Hesi horretan kokatuta dagoen ZAP proteinari esker, nukleoan ekoiztatutako RNA erretrobiralak ez da zelularen zitoplasman agertzen. ERVen adierazpena DNA metilazioaren bidez mugatu daiteke ere, DNA metiltransferasa proteinen bidez. Mekanismo hau elementu transposakorren aurkako defentsamoduan sortu zela proposatu izan da (Yoder et al., 1997). ERVen adierazpena histonetan aldaketak eginez molda daiteke, batez ere histonen metilazioaren eta azetilazioaren bidez. Adibidez, ikusi da histona metiltransferasa proteinen bidez, H3 histonaren isatsaren trimetilazioa elementu errepikakorrek isilaraztean parte har dezakeela. Azkenik, RNA interferentziaren bidez ERVen adierazpena mugatu daiteke. siRNAk (*small interfering RNA*) eta piRNAk (*Piwi-interacting RNA*), transkripzioaren mailan eta transkripzio-ostekoan LTR-eretroelementuetatik defenda dezakete genoma (Goff, 2004; Maksakova et al., 2008).



I-7. irudia: Erretrobirusen aurkako defentsahesiak.
(Goff, 2004)-tik moldatua.

1.4.1.3. Finkatzea

Genomaren eskualde baten errekonbinazioaren tasa elementu transposakorren finkatzearekin erlazioa duela proposatzen duten ereduak plazaratu dira (Duret et al., 2000; Eickbush & Furano, 2002; Biemont & Vieira, 2006). Hala ere, gizakien ERVekin (HERV) egindako azterketa batean (Katzourakis et al., 2007a) ez zuten halako erlaziorik aurkitu. Beren azterketan ikusi zuten eskualde bateko gene-dentsitatea HERV finkapenarekin erlazionatuta dagoela eta ez errekonbinazioarekin. Halaber, ikusi zuten gene-dentsitatearen eta ERVen finakatzearen arteko erlazioa korrelazio negatiboa zela.

HERVetan, ERV osoak eta solo-LTRen arteko proportzioak aztertuz, genomaren eskualde bateko errekonbinazio-tasaren eta proportzio horren arteko erlazio esangarria ikusi zen. Errekonbinazio-tasa handiagoa zuten eskualdeetan, solo-LTRen proportzioa handiagoa da, hots, eskualdearen errekonbinazio-tasak ERVen egituraren osotasuna mantentzea baldintzatzen du (Katzourakis et al., 2007a).

1.4.2. ERV kopia-kopuruaren emendioa

ERVen kopia-kopurua handitzeko, bi mekanismo nagusi daude: eretrotransposizioa eta berrinfekzioa. Eretrotransposizioan, ERVen kopuruaren emendioa hozi-lerroko zelulen barruan gertatzen da eta bi modutara gerta daiteke: *cis* eran edo *trans* erako osaketan. Berrinfekzioan, berriz, hozi-lerroko zeluletan edo zelula somatikoetan jatorria duen ERV batek hozi-lerroko zelulak infektatzen ditu. Mekanismo horietaz gain, ERV ez-funtzionalek adierazitako ERVekin, XRVekin edo genomaren beste leku batean dagoen ERV batekin errekonbina daitezke eta, batzuetan, prozesu horren ondorioz, funtzionaltasuna berreskura dezakete (Gifford & Tristem, 2003; Belshaw et al., 2004).

1.4.2.1. Eretrotransposizioa

Eretrotransposizioan, ERVak zelulatik ateratzen ez direnez, ez da beharrezkoa *env* genea funtzionala izatea. Eretrotransposizioaren *cis* eran, ERVek berek kodetutako proteinak erabiltzen dituzte ostalariaren genoman zehar mugitzeko, beraz, *gag* eta *pro-pol* geneek funtzionalak izan behar dute. (Belshaw et al., 2004). *trans* erako osaketan, berriz, eretrobirus-laguntzaile batek (endogenoa zein exogenoa) emandako proteinak erabiltzen dituzte ERVek mugitzeko. Hortaz, ERVek gene funtzionalik edukitzea ez da beharrezkoa, eskualdeak mantentzea baizik: PBS, PPT, paketatze-seinalea eta LTRak esaterako (Gifford & Tristem, 2003; Belshaw et al., 2004).

1.4.2.2. Berrinfekzioa

Berrinfekzioari dagokionez, ERV batek bere burua kopiatzen du; zelulatik ateratzen da, eta hozi-lerroko beste zelula bateko genoman txertatzen da. Hortaz, ERVak guztiz funtzionala izan behar du; hau da, gene guztiek beren funtzioa mantendu behar dute, *env* genea barne. Beraz, *env* genea gakoa da, kopia-kopuruaren emendio-mekanismoak bereizteko, eretrobirusari zelulatik irteteko eta zeluletan sartzeko gaitasuna ematen dion genea baita (Gifford & Tristem, 2003; Belshaw et al., 2004).

Berrinfekzio bidez, ERVak nola beren kopurua emendatzen duten ulertzeko, Katzourakis-ek eta beste hainbatek eboluzio-esparru bat proposatu zuten beren lanean (Katzourakis et al., 2005). Haien iritziz, ERV familia bateko kide aktibo eta ez-aktiboak denboran zehar aldatzen dira eta kopuruaren emendioa lau prozesuren arabera da:

- berrinfekzio bidez erreplikatzeko diren elementu aktiboen tasa.
- berrinfekzioan mutazio bidez inaktibo bihurtzeko diren elementuen proportzioa.
- ostalariaren zelulen banaketan mutazio bidez inaktibo bihurtzeko diren elementuen tasa.
- errekonbinazio deletereoaren bidez ezabatzen diren elementuen tasa.

ERV familia baten kide-kopurua prozesu horien arteko elkarrekintzen menpe dago, hau da, ERV berriak sortzearen (lehen bi prozesuak) eta ERVak galtzearen (azken biak) mende. Prozesu

horiek erabiliz, Kotzourakis eta besteek bi eredu eraiki zituzten ERVak beren kopurua nola emendatzen duten azaltzeko: eredu determinista eta eredu estokastikoa (Katzourakis et al., 2005).

Eredu deterministak ERVen kopia-kopuruaren emendioa aurrean dezake epe luzera. Eredu horren arabera, berrinfekzioen, inaktibo bihurtzeen eta ezabatzeen arabera, ERV familiek beren kopurua mugagabe emenda dezakete, eta ostalariaren genoma suntsitu edo taldea genomatik ezabatu. Kontrolik gabeko kopuruaren emendioan, ostalariak defentsa-sistema eraginkorrak izateko hautespen-presioa jasango luke. Familia ezabatzean, berriz, ERV aktiboak gero eta urriagoak dira eta kopuruaren emendioa gelditu egiten da. Hala ere, agertoki horretan, ERV familiek kide ugari izatera hel daitezke desagertu baino lehenago; hortaz, ERV familia handiak desagertutara bideratuta dauden eboluzio-arrastoak izan daitezke (Katzourakis et al., 2005).

Eredu estokastikoak ERV kopurua txikia denean du eragina. Eredu horretan aipatutako lau prozesuak adarkaduran gertatzen dira. Eredu horrek izar-itxura duen filogenia asimetrikoak aurreikusten ditu. Halako filogeniak hainbat ERV familiatan agertzen dira. Zuhaitz topologia horiek aditzera eman dezakete ERVen kopuruaren emendio-tasa denboraren puntu batean izugarri aldatu zela, adibidez, ostalariaren defentsa mekanismoen aldaketen ondorioz (Katzourakis et al., 2005).

I.4.2.3. Gizakien ERV familien emendioa

ERVen kopuruaren emendioa, batez ere, gizakien ERVetan (HERV, *human endogenous retrovirus*) aztertu da (**1. taula**). Oro har, ikusi da HERV familia gehienek berrinfekzio bidez beren kopurua emendatu dutela (Belshaw et al., 2004; Belshaw et al., 2005). Ondorio hori HERV familia gehienetan *env* genearen dn/ds proportzioa (aldaketa ez-sinonimo eta sinonimo proportzioa) 1 baino txikiagoa izatean oinarritzen da, hau da, *env* genean hautespen purifikatzailea gertatzen dela, beraren funtzioa mantentzeko presioa dagoela. Gorago azaldu bezala, berrinfekzioa gertatzeko, behar-beharrezkoa da *env* geneak funtzionala izaten jarraitzea. Are gehiago, familia batzuetan proportzio hori XRVetan gertatzen den proportzioetara hel daiteke, HERV-T (I klasekoa) edo HERV-S (III klasekoa) familietan gertatzen dena, esate baterako (Belshaw et al., 2004; Belshaw et al., 2005).

I-2. taula: HERV familien kopia-kopuruaren emendioaren mekanismo nagusiak.

Atal honetan azaldutako adibideetan oinarritua.

HERV familia	Berrinfekzioa	Erretrotransposizioa <i>cis</i>	Erretrotransposizioa <i>trans</i>
HERV-T	X		
HERV-S	X		
HERV-K(HML2)	X		
HERV-K(HML3)		X	
HERV-H			X
ERV9			X
HERV-W			X

II klaseko HERV-K(HML2) familian azterketa sakonagoak egin dira (Belshaw et al., 2004). Familia interesgarria da, gizakiengan orain dela gutxi arte aktibitatea eta kopuruaren emendioa izan dituen HERV familia baita. Ikusi da familia horretan *trans* erako osaketa ez dela ohikoa, gene errotobiral nagusietan stop kodoiak, oro har, ez baitira mantendu. Hala ere, familia horretako kideren batean, jatorri ezberdina duten geneak aurkitu dira. dn/ds proportzioari dagokionez, *env* genean 1 baino txikiagoa izateaz gain gainontzeko geneetan, balio hori agertzen da. dn/ds proportzioa barne-adarretan kanpo-adarretan baino txikiagoa da, eta, hala, ERVek egitura mantendu izanaren ideia indartu egiten da eta, hori dela-eta, ondorioztatu da HERV-K(HML2) familian berrinfekzioa kopuruaren emendio mekanismo nagusia izan dela (Costas, 2001; Belshaw et al., 2004).

Hala ere, ikusi da kopia-kopuru handienak dituzten familiek beren kopia-kopurua erretrotransposizioaren bidez emendatu dutela (Belshaw et al., 2005). Halaber, ikusi da HERV-H familiako klado batean eta ERV9 familian emendioa *trans* erako osaketan gertatu dela, *pol* eta *env* geneen dn/ds proportzioa handia baita. HERV-K(HML3) familian, berriz, ondorioztatzen da erretrotransposizioa *cis* erara gertatu dela, *env* genearen dn/ds proportzioa handia delako baina *pol* genean, txikia. Azkenik, HERV-W familiaren kideen bi herenen LINE erretrotransposoiaren bidez emendatu dute beren kopurua (Belshaw et al., 2005).

1.4.3. ERVen biziraupena

1.4.3.1. Mutazioa

ERVek beren egitura (LTR-*gag-pro-pol-env*-LTR) mantentzen badute ere, oro har, mutazio-metaketa ondoz geneak ez-funtzional bilakatzen dira eta ez dute proteinarik kodetzen. LTRak, geneak ez bezala, aktibo mantentzen daitezke edota genomak bakarrik agertu, solo-LTRak deitzen diren egitura gisa agertuz (Griffiths, 2001; Belshaw et al., 2005).

Oro har, erretrobirusen genomatik mutazioen eragin gutxien jasotzen duen eskualdea *pol* geneko alderantzizko transkriptasa (RT) eskualdea da. Eskualde hori ordeztzen ez-sinonimoetatik nahiko babestuta dago; hainbat motatako erretrobirusek eskualde hori nahiko ondo kontserbatuta dute eta, hortaz, ERVak identifikatzeko erabiliena da. Eskualde horretan oinarritzen dira erretrobirusen filogenia ikertzen duten lan gehienak. Hala ere, badaude *env* geneko transmintza (TM) eskualdean oinarritzen diren lanak, hainbat birusetan uste zen baino kontserbatuagoa baitago (Benit et al., 2001; Gifford & Tristem, 2003).

ERVak genomatik txertatzerakoan LTRak berdina dira. Denboraren poderioz, LTRak elkarrengandik ezberdintzen dira, DNA erreplikazioaren ondoz jasotzen dituen ordeztzen, txertatze eta ezabatzeen eraginez. LTR arteko dibergentzia baliagarria da ERVen txertaketa-denbora estimatzeko, hots, erloju molekular moduan erabili daiteke. HERVen LTRekin egindako analisian, LTRak independenteki eboluzionatu dutela eta berrantolaketa gertaerarik egon ez dela onartuz, ikusi zen hainbat HERVen LTRen ordeztzen-tasa $2,3 - 5,0 \times 10^{-9}$ ordeztzen/lekuko/urteko dela. Ordeztzen-tasa hori kalibratzeko mundu zaharreko tximinoen eta tximinoen arteko banaketa erabili zen eta ordeztzen-tasa hau mamalioen pseudogene eta geneen eskualde ez-kodetzaileen antzeko tasa da (Dangel et al., 1995; Johnson & Coffin, 1999).

1.4.3.2. Errekonbinazioa

ERVetan, errekonbinazioa bi mailatan gerta daiteke: gene bakoitzaren barruan eta geneen artean.

Gene arteko errekonbinazioa

Horago aipatu bezala, RT eskualdean eta TM eskualdean oinarritutako filogeniak egin dira eta horiek hainbat ezberdintasun dituzte. Gertaera hori azaltzeko proposatu da zati zentralako geneen artean errekonbinazioa dagoela, hainbat elementutako RT eta TM eskualdeak jatorri ezberdina dutela ikusita. Halaber, lehen aipatu bezala, HERV-K(HML-2) familiako hainbat kidek egitura kimerikoa azaltzen dute; jatorri ezberdina duten LTRez edo geneez osatuta daude. Azkenik, eskualde zentrala eta LTR arteko errekonbinazioa ere detektatu da, jatorri ezberdina duten LTR eta eskualde zentralak ERV gutxi batzuetan batera agertzen baitira (Benit et al., 2001; Costas, 2001; Belshaw et al., 2004; Hughes & Coffin, 2005).

Gene barruko errekonbinazioa

Gizakiaren HERV-V familiaren *env* genea aztertzen zen lan batean (Kjeldbjerg et al., 2008) ikusi zen gene horretan gene-konbertsioa gertatu zela. Lan horretan bata besteari atzean dauden HERV-Vko kideen *env* geneak hainbat tximinotan aztertu zituzten eta espezie guztietan bi *env* gene horien arteko gene-konbertsioa detektatu zuten; hots, gene horien eskualde bat gene batetik bestera transferitu zela (Kjeldbjerg et al., 2008). Txerriaren ERVen *env* genean ere, *env* mota ezberdinen artean, PERV-A eta PERV-B motak, alegia, errekonbinazio-gertaerak detektatu dira (Lee et al., 2008). Azkenik, ikusi da gizakiaren HERV-K(HML2) familiako ERVen LTRetan errekonbinazio edota gene-konbertsio gertaerak nahiko ohikoak direla, alde batetik, familia horretako hainbat kidek historia filogenetiko ezberdina zutelako eta, bestetik, familia horretako kide bat gene-konbertsioaren bidez ezabatu zelako (Hughes & Coffin, 2005).

1.4.3.3. Txertatze-norabidea

ERVen txertatze-norabidea aztertzean ikusi da gizakien geneen introietan agertzen diren ERVek kontrako norabidean txertatzeko joera dutela (Medstrand et al., 2002). Joera hori, elementu

transposakorretan ere detektatu da gizakiengan eta saguengan (van de Lagemaat et al., 2003) baita behien LTR erretrotransposoietan ere (Almeida et al., 2007). Joera hori introietan ERVen biziraupenarekin erlazionatu da (van de Lagemaat et al., 2006). Gertaera horren azalpen gisa proposatu da norabide bereko txertaketak, LTRen poliadenilazio-seinalea dela-eta, transkripzioa goiz bukatzea eragiten duela. Horrek norabide bereko txertaketak kontrako hautespena jasotzea eragingo luke (van de Lagemaat et al., 2006).

1.4.4. Genoma-inbasioen adibideak

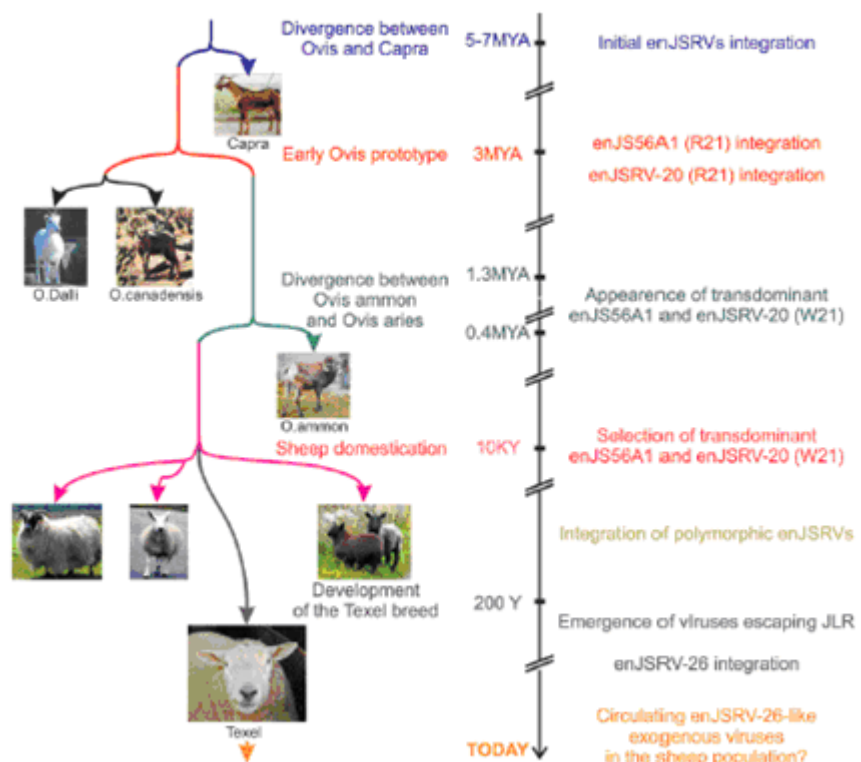
ERV familia bat sortzeko prozesua - hau da, erretrovirus batek infekzio sortzen duenetik ERV batek populazio-ostalariaren genoma kolonizatu arte -, populazio basatietan aztertzea zaila suertatu bada ere, koalaren erretrovirusari (KoRV) esker, prozesu hori zuzenean aztertu ahal izan da (Tarlinton et al., 2006; Tarlinton et al., 2008).

Koala Australiako ekialdean bizi den marsupial bat da eta haren genomak finkatzen ari den ERV bat, KoRVa, hain zuzen ere, hauteman da. KoRV ERVak giboien GALV XRVarekin erlazio filogenetiko estua duela ikusi da, eta horiek *Gammaretrovirus* (I klasea) generoaren barruan kokatzen dira. Erlazio hori dela-eta uste da hirugarren espezie baten bidezko espezie arteko saltoz gertatu dela. Espezie bitartekari hori (edo horiek) Asiako karraskariak izan daitezkeela proposatu da, GALV XRVarekin erlazionatutako ERVak detektatu baitira espezie horietan (Tarlinton et al., 2008).

KoRVa koalen neoplasia hematopoietiko gaixotasunarekin erlazionatu da eta adierazpen-maila altua du. Ezaugarri horiek XRV batenak badira ere, argi dago KoRVa oraindik finkatu gabeko ERV bat dela, hozi-lerroko genomak txertatuak baitaude eta indibiduo artean kopia kopurua eta sekuentzia erretrobirala aldatokorrak baitira. Gainera, ikusi da txertatutako KoRV gehienak egituraren osotasuna mantentzen dutela, koala batzuetan kopia ez-funtzionalak aurkitu badira ere. Bestalde, badaude KoRV gabeko populazioak, gainontzeko



I-8. irudia: KoRVren agerpena koalaren hainbat populaziotan (Tarlinton et al., 2006).



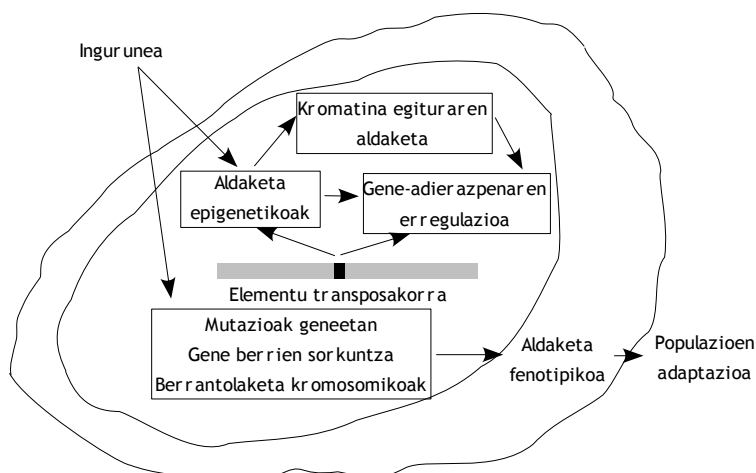
I-9. irudia: enJSRVen eta ardiaren historia ebolutiboa. (Arnaud et al., 2008).

populazioetatik isolatuta dagoen populazioa bat, alegia, eta, beste batzuetan, agerpena aldakorra da. Aldakortasun horrek aditzera ematen du infekzioa populazioan sakabanatzen ari dela, Australiako iparraldetik hegoaldera (I-8. Irudia). Uste da une honetan oraindik gertatzen ari den prozesu hori orain dela 200 bat urte hasi zela (Tarlinton et al., 2006; Tarlinton et al., 2008).

Koalaren adibideaz gain, ardiaren genomak, enJSRV birusaren inbasioa dugu (I-9. irudia) (Arnaud et al., 2008). enJSRVa *Jaagsiekte sheep retrovirus* erretrobirus exogenoaren bertsio endogenoa da. Ardiaren genoma *Ovis* (ardiak) eta *Capra* (ahuntzak) generoak banandu zirenean (orain dela 5-7 miloi urte) inbaditzen hasi zela jo da. Inbasioa hasi zenetik, enJSRVren kopiek eboluzionatzen eta inbaditzen jarraitu dute ardiaren genometan. Horren ondorioz, enJSRV kopia batzuk *Ovis* generoko hainbat espezieetan daude eta beste batzuk, ardiaren soilik. Gainera, ardiaren kopia batzuk polimorfikoak dira, hots, bakarrik ardiaren proportzio batean agertzen dira. Halaber, ondorioztatu da kopia horiek ardiaren espezieaz gertatu baino beranduago (orain dela 10.000 urte) txertatu direla. Hortaz gain, jo da horietako kopia bat, LTRen dibergentzia oinarrituta, orain dela 200 urte txertatu zela. Proposatuta da gaur egun enJSRVk ardiaren genoma inbaditzen jarraitzen duela (Arnaud et al., 2008).

I.5. ERVen adierazpena eta ostalariaren duten eragina

Erretrobirus endogenoak (ERV), elementu transposakorrak diren heinean (TE), TEek ostalarietan duten eraginak dituzte (10. irudia). TEek, geneetan, mutazioak eragin ditzakete; gene berriak sorkuntzen lagundu dezakete edota berrantolaketa kromosomikoetan bitartekaria izan daitezke (Biemont & Vieira, 2006). Berrikuntza-iturria izateaz gain, aldaketa epigenetikoaren eraginean ondorioz, TEak adieraz daitezke eta geneen adierazpena erregulatu. Eragin horien ondorioz aldaketa fenotipikoak gerta daitezke, eta populazioari adaptazio-bide berriak eman (Biemont & Vieira, 2006).



I-10. irudia: Elementu transposakorrak eta genomaren ingurune elkarrekintzak.

(Biemont & Vieira, 2006)-tik moldatua.

Ostalariaren genomak TEek egindako ekarpena ezberdina da eskualdearen arabera. Oro har, TEek ekarpena txikia egin dute adierazten diren eskualdeetan; gizakiengan kodetzen diren proteinen %4ean, TEen sekuentziaren bat detektatu zen (Nekrutenko & Li, 2001); eta LTR erretrotransposoiek saguei egindako ekarpena %1,5 da (DeBarry et al., 2006), *Drosophilari* %1,9koa (Ganko et al., 2006) eta *C. Elegansari* %0,6koa (Ganko et al., 2003). Hala ere, UTR eskualdeetan, TE ugari agertzen dira, %27 gizakiengan eta %18 saguengan (van de Lagemaat et al., 2003), eta eskualde sustatzaileetan, %25

gizakiengan (Jordan et al., 2003). Lehenago aipatu bezala, ERVak TEen barruan sailkatzen dira eta, hortaz, ERVek ekarpenak egin diezazkiokete ostalariaren genomari.

Aurreko atalean ERVek populazioetan izan ditzaketen eraginak aurkeztu ditugu eta, atal honetan, ostalariarekin dituzten elkarrekintzak eta ERVen adierazpena ikusiko ditugu.

I.5.1. ERVen eta ostalariaren arteko elkarrekintza

ERVak ostalariaren geneetatik hurbil txertatzen badira, beren LTRek gene horien adierazpena alda dezakete. Adibidez, gizakiengan, HERV-E familiako kide bat txertatzeak listuan amilasa adieraztea eragin du (Samuelson et al., 1990). Halaber, jatorrizko transkripzioa areagotu dezakete, genetik distantzia-hein zabal batean txertatuz; eragina ERVtik 100kb-ra heldu daiteke (Jern & Coffin,

2008). Gainera, solo-LTR-ek ere sustatzaileetan eta areagotzaileetan eragin dezakete, eta ostalariaren geneetan eragin handia izan. ERVek geneen adierazpena aldatzeko beste modu bat genomotitsasketa alternatiboak sortzea da. Azkenik, ERVek genomaren antolaketan eragin dezakete antzeko ERV arteko errekonbinazioaren bidez (Jern & Coffin, 2008):

- LTR arteko errekonbinazioaren eraginez solo-LTR-ak gelditzea.
- Kromosoma eta norabide berean dauden bi ERVen arteko errekonbinazioa gertatzea, eta birusak eta kromosoma zati bat galtzea.
- Kromatida ezberdinetan dauden ERVen 5' eta 3' LTRen arteko errekonbinazioa, eta kromatida batean ERVaren zati zentrala tandemean geratzea eta bestean solo-LTRa.
- Gene-konbertsioa gertatzea, eta ERV baten sekuentzia zati bat beste ERVan kopiatzea.

Prozesu horiek direla-eta, sinesgarria izan daiteke ERV arteko errekonbinazio gertaerek eskualde genomikoen, exoien eta kontrol-eskualdeen nahasketan eragina izan dutela, eta, hala, horiek beste testuinguru batera eramango ditu, non, genomaren dinamikan eragin handia duten (Jern & Coffin, 2008).

1.5.1.1. Eraginak

ERVek, ostalarian, eragin ezberdinak izan ditzakete. ERV familiaren arabera, eragin horiek onuragarriak izan daitezke, adibidez, ostalariari funtzio erabilgarriak emanez, edo kaltegarriak, gaixotasunetan zerikusia izanda.

ERVen eragin onuragarriak ikertuena gizakiaren sinizitina proteinak dira. Proteina horiek ERV ezberdinen *env* genearen produktuak dira eta, placentaren garapenean, zeregin garrantzitsua dute. Sinizitina-1 deritzon proteina HERV-W familiako *env* genetik eratortzen da. Proteina hori placentaren kanpoaldea eratzen duten tropoblastoetan adierazten da eta zelulen fusioan eta sinizitinoaren eraketan hartzen du parte. Sinizitina-2 proteina ere hauteman da; horrek sinizitina-1 proteinaren ezaugarri berdinak ditu eta HERV-FRD familiaren *env* genea da (Jern & Coffin, 2008).

Antzeko funtzioa eta jatorri erretrobirala duten proteinak saguan detektatu dira, sinizitina-A eta sinizitina-B eta saguaren ERVen *env* genearen produktuak dira. Ardianm JSRV birusaren forma endogenoaren *env* geneak ere antzeko ezaugarriak ditu. Gene horiek jatorri dituzten ERVek ez dute ahaidetasun filogenetikorik eta espezie bakoitzak modu independentean lortu du; hau da, eboluzioan konbergentzia bat gertatu da (Jern & Coffin, 2008).

Beste eragin onuragarri bat saguetan gertatzen da. Saguetan MMTV birusaren forma endogenoek eta exogenoek superantigeno bat adierazten dute, zeinak mota ezberdinetako birus exogenoak modu eraginkorrean blokeatzen dituen (Jern & Coffin, 2008).

ERVen eta hainbat gaixotasunen arteko erlazioari buruzko eztabaida dago; ERVak hainbat gaixotasun neurodegeneratibo eta minbiziren jatorria edo ondorio ote diren eztabaidatzen da, hain zuzen ere. Jern eta beste hainbaten (2008) ustez, ERVak gaixotasunekin lotzeko saiakerek oinarri genetiko nahikorik gabeko asoziazioak dira eta, ondorioz, gehiegi interpretatutako ondorioak lortzen dira. (Jern & Coffin, 2008). Hala ere, hainbat gaixotasunetan ERVen funtzioa edo eragin-mekanismoa argi ez badago ere, ERVen adierazpena hainbat gaixotasunetan detektatu da (Jern & Coffin, 2008). Adibiderik argienak saguetan aurkitu dira. MLVaren eta MMTVaren forma endogenoak eta IAP partikulak gaixotasun genetiko eta minbiziekin erlazionatu dira. Gizakiengan, aldiz, ez da, oro har, ERV berrien adierazpen-erlazio argirik aurkitu. (Jern & Coffin, 2008). Hala ere, gizakiengan, ikusi da esklerosi multiplean HERV-W familiaren *env* genea adierazten dela eta gene horrek erantzun immunea abiarazten duela (Perron et al., 2009). Familia horren *env* genearen zergatik adierazten den guztiz argi ez badago ere, ikusi da Herpesvirusak egonez gero aktiba daitekeela, hots, aktibazioa ingurumenaren arabera izan daitekeela (Perron et al., 2009).

Hala ere, gizakiengan ikusi da ERVen adierazpena aztertutako minbizidun ehunetan gainerregulatuta dagoela. Adibidez, lesio aitzindaridun eta haurdunaldi jatorridun kartzinometan

ikusi da testikuluko eta obarioko hozi-zelulako tumoreetan HERV-K familia adierazten dela. Gainera ikusi da familia horren *gag* eta *env* geneen adierazpenaren eta terapiaren porrotaren eta metastasiaren artean korrelazioa dagoela (Herbst et al., 1996). Halaber, ikusi da zelula hematopoietiko gaiztoetan ere HERV-E familiaren kide batzuen eta erlazionatutako ERVen *gag* genearen adierazten dela (Prusty et al., 2008) edota badagoela HERV-Kren *pol* geneko RT eskualdearen adierazpenaren eta bularreko minbiziaren pronostikoaren arteko erlazioa (Golan et al., 2008). Oraindik ere, gaixotasun horietan ERVek zerikusirik ote duten argitzeke dago; lehen aipatu bezala, jatorria edo ondorio diren zehaztu gabe dago (Jern & Coffin, 2008).

Gizakiengan ere ERVak esklerosi multiplearekin eta eskizofreniarekin erlazionatzen saiatu dira. Ikerketek aditzera eman dute HERV-W eta HERV-H familiak bi gaixotasun horietan gainadierazten direla, baina ez dago argi zer-nolako zerikusia duten (Jern & Coffin, 2008).

1.5.2. ERVen eta XRVen arteko elkarrekintza

Erretrovirus endogenoen eta exogenoen arteko ahaidetasuna dela-eta, proposatu da haien arteko elkarrekintza egon daitekeela, batez ere ostalaria defendatzearekin erlazionatuta. Hala ere, adibide gutxi batzuek elkarrekintza hori frogatu dute, besteak beste oiloan eta ardian.

ERVen eta XRVen arteko interferentziaren lehen adibide argia oiloetan aurkitu zen (Denesvre et al., 2003). ev/J 4.1 deritzon ERVaren *env* genea erabiliz, ALV-J XRVa zelulan ez sartzea lortu zen; *env* geneen arteko interferentzia zegoela ondorioztatu zuten.

Ardiaren JSRV birusa ERVen, XRVen eta ostalarien arteko koeboluzioaren adibidetzat jarri izan da (Arnaud et al., 2008). JSRV erretrovirusak forma endogenoa (enJSRV) eta exogenoa (JSRV) ditu. enJSRVak ardiaren *conceptus*aren eta plazentaren morfogenesian hartzen du parte, prozesu horretan parte hartzen duten hainbat zelulatan adierazpena ikusi baita, batez ere tropoblaston eta sinzitioan, eta horiek *conceptusa* elikatzeko erabiltzen diren plazentomak ezartzeko eta eraketzeko beharrezkoak dira. Zelula horietan enJSRVaren *env* genea ez adieraztea lortzen bada, haurdunaldia eteten da. Halaber, ikusi da fetuan T-linfozitoen eraketan parte hartzen duten zeluletan ere enJSRVaren *env* genea adierazten dela, horrek ardiari *Betaretrovirus*en aurreko nolabaiteko tolerantzia ematen dio (Spencer et al., 2003; Arnaud et al., 2008).

enJSRVak ere JSRVarekin bi puntutan interferitzen duela ikusi da: JSRVaren sarreran eta erreplikazio-zikloaren bukaeran. enJSRVaren adierazpen altua duten zeluletan, JSRVaren sarrera oztopatuta dago, ziurrenik Hyal-2 hartzailearen interferentziaren ondorioz. Hori hartzailea saturatzen duelako gerta daiteke edo hartzailearean gainazaleko adierazpena murrizten duelako. Erreplikazio-zikloaren bukaeran, enJSRVaren Gag proteinak JSRVaren Gag proteinekin lotzen dira eta partikula kimerikoak sortu. Partikula horiek zelularen mintzera garraiatu ezin daitezkeenez, degradatu egiten dira (Arnaud et al., 2008).

1.5.3. ERVen adierazpena

Goiago aipatu bezala, ERVek, gaixotasun batzuetan, zerikusia izan dezakete. Hori dela-eta, ERVen adierazpena aztertu da, gizakiengan batez ere. Metodo ezberdinen bidez, ERV familiek duten adierazpena aztertu da, bai ehun bakoitzean, bai garapen fase bakoitzean edo gaixotasunetan. Lan horretan ikusi da ERV familia bakoitzaren adierazpena modu ezberdinean gertatzen dela.

1.5.3.1. Gizakien ERVen adierazpena

Azterketa batean ikusi zen HERV-R eta HERV-K familiak, *in situ* hibridazioaren bidez, fetuaren garapenean adierazten zirela eta haien *env* geneak hainbat ehunetan adierazten zirela. HERV-R familiaren *env* genea gizakien enbriogenesian erregulatuta zegoela ondorioztatu zuten, guruin adrenalean adierazpen nahiko altua baitzuen. Halaber, Northern blot-en bidez egindako azterketetan ikusi zen HERV-F, HERV-H, HERV-K(HML-6), HERV-W (sinzitina kodetzen duen ERVa) eta HERV-FRD (sinzitina2 kodetzen duena) familiak plazentan adierazten direla. Gainera, familia horiek garapenean zehar adierazpen-aldakortasunak azaldu zuten: konstante mantentzen zirenak (HERV-V

eta HERV-F), denborarekin gutxitzen zirenak (HERV-FRD) eta handitze esangarria zutenak (HERV-H eta HERK(HML-6)) (Andersson et al., 2002; Okahara et al., 2004).

Gizakien ERVen adierazpena sakonago aztertzeke, DNA-mikromatrise bat sortu zen (Seifarth et al., 2005) eta ehun osasuntsuetan ez ezik, mikromatrise hori eskizofreniadun eta desoreka-bipolardun pertsonen ehunetan (Frank et al., 2005), bular-minbizidun ehunetan (Frank et al., 2008) eta mundu zaharreko tximinoen ehunetan (Stengel et al., 2006) ere erabili da. Lan horietan, behin ERV familia interesgarriak detektatu eta gero, RT-PCR bidez familia horien adierazpen-maila aztertu zen (I-3. taula).

I-3. taula: Mikromatriseen bidez egindako ERVen adierazpenari buruzko lanen laburpena.

X, adierazpena detektatu zenean; * adierazpenean ezberdintasun esangarria izan zenean; -, daturik ez; (Frank et al., 2005; Seifarth et al., 2005; Stengel et al., 2006; Frank et al., 2008)-tik moldatua.

		I. klasea (HERV-)										II. klasea (HML-)										III. klasea
Ehuna		I	T	FRD	E	H	F	W	R	ERV-9	1	2	3	4	5	6	7	8	10	HERV-L		
Osasuntsuak																						
Odola			X	X	X		X	X		X		X	X	X		X	-		X			
Garuna	Gizakia			X	X		X	X		X		X	X	X		X	-		X			
	Makakoa	-	-		X	-			-		-		X	X	-		-	-	X			
	Mandrila	-	-			-			-	X	-				-		-	-	X			
Azala		X	X	X	X	X	X	X	X	X		X	X	X	X	X	-		X	X		
Muskulu eskeletikoa	Gizakia	X	X	X	X	X	X	X	X	X			X	X	X	X	-			X		
	Makakoa	-	-		X	-		X	-	X	-		X		-	X	-	-	X			
	Mandrila	-	-		X	-	X	X	-	X	-		X		-		-	-	X			
Bihotza		X	X	X	X	X	X	X	X	X			X	X		X	-			X		
	Gizakia				X	-		X	-	X	-		X		-		-	-	X			
	Makakoa	-	-		X	-		X	-	X	-		X		-		-	-	X			
	Mandrila	-	-		X	-	X	X	-	X	-		X		-		-	-	X			
Giltzurruna				X	X		X	X	X	X		X	X	X		X	-	X				
	Gizakia				X	-		X	-	X	-		X	X	-		-	-	X			
	Makakoa	-	-		X	-		X	-	X	-		X	X	-		-	-	X			
	Mandrila	-	-		X	-		X	-	X	-				-		-	-	X			
				X	X		X	X	X	X		X	X	X		X	-					
Gibela				X	X		X	X	X	X	-		X	X	-		-	-	X			
	Makakoa	-	-		X	-		X	-	X	-		X	X	-		-	-	X			
	Mandrila	-	-		X	-	X	X	-	X	-				-		-	-	X			
	Gizakia			X	X		X	X	X	X		X	X	X		X	-					
Biririkak					X	-		X	-		-		X	X	-		-	-	X			
	Makakoa	-	-		X	-		X	-		-		X	X	-		-	-	X			
	Mandrila	-	-		X	-		X	-		-		X	X	-		-	-	X			
Urdaila					X		X	X		X			X	X		X	-					
Kolona				X	X		X	X		X		X	X	X		X	-					
Ondestea				X		X				X		X	X	X		X	-					
Bularra				X	X		X	X		X		X	X	X	X	X	-	X				
Obarioa				X		X				X		X	X	X		X	-					
Uteroa		X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	-	X	X	X		
Utero-lepoa		X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	-		X	X		
Plazenta		X	X	X	X		X	X	X	X		X	X	X	X	X	-		X	X		
Testikulua		X	X	X	X		X	X	X	X		X	X	X	X	X	-		X	X		
Prostata					X		X	X		X	X	X	X	X	X	X	-	X	X			
Tiroidea		X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	-	X	X		
Eskizofrenikoa vs bipolarra		*															*			*		
Eskizofrenikoa vs osasuntsua												*		*			*					
Bipolarra vs osasuntsua										*		*					*					
Bular minbizia vs osasuntsua				*	*					*		*				*	-					

Mikromatrise horrekin, gizakiaren 20 ERV familiaren *pol* genearen aktibitatea aztertu zen 19 ehunen RNAn. Familiaz familia eta ehunez ehun, familia horien aktibitatea aldatu egiten zen. I. klasekoen ERV artean, ERV9 familia ehun guztietan detektatu zen; HERV-E, HERV-F, HERV-W eta HERV-FRD ehun gehienetan detektatu ziren eta HERV-H (genoman kopia kopuru handiena duen familia, gorago aipatu bezala) HERV-I eta HERV-T familiak ehun gutxi batzuetan adierazten dira. II. klasean, HML-3 eta HERV-K(HML6) familien adierazpena ehun guztietan detektatu zen, HML-4 ehun gehienetan eta HERV-K(HML-2) familia ehun mota gutxitan. III. klaseko familiak, HERV-Lk, gainontzeko klasekoak baino adierazpen gutxiago du (Seifarth et al., 2005).

Mikromatrise berdina erabilita, aztertu zen HERV familiak nola adierazten ziren eskizofreniadun eta desoreka-bipolardun pertsonen ehunetan. Hainbat HERV familia, egoera normalean, garunean adierazten dira. HERV-I, ERV9, HML-7 eta HERV-L familien kasuan gaixoen eta osasuntsuen

artean adierazpen-ezberdintasun batzuk aurkitu badira ere, ez dira esangarriak. HERV-K(HML-2) familian, berriz, ikusi zen adierazpen-ezberdintasuna esangarria zela (Frank et al., 2005).

Mikromatrize hori bular-ehun osasuntsu eta minbizidunen arteko HERV familien adierazpen-ezberdintasuna aztertzeko ere erabili zen. Oro har, ehun osasuntsuek eta minbizidunek antzeko HERVen adierazpen-profila zuten, baina, minbizidun ehunetan, HERVen adierazpena murriztuta zegoen. Hala ere, aztertutako paziente multzo batean, HERV-T, HERV-E, HERV-F (I. klasea) eta HERV-K (II. klasea) familietako hainbat kide gainerregulatuta zeuden (Frank et al., 2008).

Azkenik, mikromatrize hori makakoen eta mandrilen ehunetan erabili zen gizakien ERVen adierazpena aztertzeko. HERV-E (garunean), ERV-9 (biriketan, garunean), HERV-FRD (aztertutako ehun guztietan), HERV-F (giltzurrunetan, biriketan), HML2 (ehun guztietan) eta HML-3 (garunean) ez ziren makakoan eta mandrilean adierazi. Lan horretan, batez ere, garunean HERVek duten adierazpen-maila aztertu zen. Oro har, ikusi zuten gizakiengan HERV familia gehiago adierazten direla eta gainerregulatuta daudela (Stengel et al., 2006).

Gizakien ERVen adierazpena aztertzeko beste hurbilketa batean, gizakien EST datu-baseak aztertu ziren gizakien 4 ERV familiaren adierazpen-maila detektatzeko. ERV familia bakoitzak adierazpen-maila ezberdina azaldu zuten ehun bakoitzean. HERV-K, HERV-H eta HERV-W familiek antzeko adierazpen maila dutela ikusi zen, HERV-E familian, berriz, baxuagoa. Metodo beraren bidez aztertu zen familia horiek minbizi ehunetan nola adierazten ziren. HERV-K familia hainbat minbizi-ehunetan (testikulueta, bularreko, eta buru-lepoko ehunetan) ehun normaletan baino maizago adierazi zen; HERV-H familia, prostata- eta kolon-ehunetan; eta HERV-E eta HERV-W familiek ez zuten ezberdintasun esangarriarik azaldu (Stauffer et al., 2004).

I.5.3.2. Txerrien ERVen adierazpena

Txerrien ERVak (PERV, *porcine endogenous retrovirus*) eta haien adierazpena aztertu dira PERVek xenotransplanteetan izan dezaketen eraginagatik. PERV familiarik aztertutakoa PERV- γ 1 familia izan da, batez ere PERV-A, PERV-B eta PERV-C *env* genearen motak, PERV-A eta PERV-B *env* motek giza zeluletan sartzeko gaitasuna azaltzen baitute. Familia hori hainbat txerri-arraza eta ehunetan adierazten dela ikusi da (Wilson, 2008).

Hala ere, PERV- γ 2 familiak ehun bakoitzean zer-nolako adierazpen-maila duen bakarrik aztertu da sistematikoki. Horretarako, familia horren *pro-pol* genearen adierazpena hainbat txerri-ehunetan aztertu zen RT-PCR bidez. Adierazpen handienak gibelean, giltzurrunetan eta barean detektatu ziren (Klymiuk et al., 2006).

I.6. Aztertutako taxoiak eta espezieak

Lan honetan erretrovirus endogenoak (ERV) bi taxoi ordezkatzeko dituzten espezieetan aztertuko dira: behia (*Bos taurus*), Ruminantia subordenaren, hots, hausnarkarien, ordezkari moduan eta zaldia (*Equus caballus*), Equidae familiaren ordezkari moduan. Atal honetan, taxoi eta espezie horiek aztertuko ditugu eta taxoi eta espezie horietako ERVei buruz dakiguna.

I.6.1. Hausnarkariak

Ruminantia subordenean (hausnarkariak), 6 familietan (Tragulidae, Giraffidae, Antilocapridae, Moschidae, Cervidae eta Bovidae) banatzen diren 200 espezie inguru sailkatzen dira. Animalia horiek kontinente gehienetan aurkitu dira eta haien erregistro fosilek aditzera eman dute orain dela 50 bat milioi urte agertu zirela. Bestalde, talde horretako animaliek dibertsitate handia dute ekologiar, jokaerari eta fisiologiar dagokionez eta, askotan, kraneoko apendizteen moduko ezaugarri bereziak, baita gorputz-tamainako aldakortasun handia ere (Fernandez & Vrba, 2005).

I.6.1.1. Behia

Behia interes ekonomiko handia duen animalia da, bai gizateriarentzat elikatze-iturria delako, bai pertsona askoren bizimoduaren oinarria delako. Hausnarkarien artean behiarena da sekuentziatutako

lehen genoma. Behiaren genomaren bi bertsio eskuragarri daude, Btau3.1 eta Btau4.0 izenekoak, BAC eta WGS metodoen konbinazioz egin dira. Genomak 2.87Gbp-ko luzera duela zenbatetsi da, eta akrozentrikoak diren 29 kromosoma pare autosomikotan eta bi kromosoma sexualetan antolatuta dagoela. Hala ere, oraindik Y kromosoma sekuentziatu gabe dago (Elsik et al., 2009).

1.6.1.2. ERVak hausnarkarietan

Hausnarkarietan egindako ERVei buruzko lan gehienak esperimentalak izan dira. II klaseko ERVak hainbat ungaratutan nola agertzen diren Southern-blot bidez aztertu zen (Hecht et al., 1996). Caprinae subfamilian hibridazioak argiak izan ziren, ardi eta ahuntz artean eredu-ezberdintasunak ikusi baziren ere. Bovidae familiako gainontzeko espezieetan ere hibridazioak gertatu ziren, baina ahulagoak (Hecht et al., 1996). Lan horretaz gain, ERVen karakterizazio sakonagoa ardietan egin da (Klymiuk et al., 2003) eta, berriki, behietan (Xiao et al., 2008b).

Ardien genomaren egon daitezkeen ERVak bilatzeko egindako ikerketak esperimentalak izan dira. ERVen detekzioa aurrera eramateko, *pro-pol* eskualde batekin lotzen diren hasle endekatuak erabili ziren. Modu horretan lortutako klonak, %90eko antzekotasuna izanez gero, talde (edo familia) batean sailkatzen ziren, txerrietan erabilitako izendapenari jarraituz eta akastun gutxien zituen klona prototipotzat hartuz. Horrela, ardian, I. klaseko 9 familia definitu ziren ($\gamma 1$ - $\gamma 9$ taldeak) eta II. klaseko 3 familia ($\beta 1$ - $\beta 3$). Familia horien kopia kopurua aldakorra da familia batetik bestera; $\gamma 2$ taldea ugariena izan zen. Halaber, ikusi zen $\beta 1$ familiako kideek ardien JSRV birusaren forma endogenoekin erlazio filogenetiko zutela (Klymiuk et al., 2003).

ERVak behiaren genomaren nola agertzen ziren, orain gutxi arte, ez zen oso ezaguna. Egindako lan esperimentalak ardietan detektatutako ERVetan eta haien sailkapenean oinarritzen dira. Horrela, behian, 4 familia definitu ziren: 3 I. klasekoak ($\gamma 4$, $\gamma 7$ eta $\gamma 9$ familiak) eta 1 II. klasekoa ($\beta 3$ familia). Bi familia horien genomak karakterizatun dira: $\beta 3$ eta $\gamma 4$ familietako ERVak. (Xiao et al., 2008a; Xiao et al., 2008b; Xiao et al., 2008c)

$\beta 3$ familiako ERV prototipoaren genomak 7.666 base-pare ditu eta eretrobirusen ohiko egitura du, *gag* (507 aminoazidoko luzera), *pro* (271 aa), *pol* (879 aa) eta *env* (603 aa) geneak eta, alboetan, LTRak (406 eta 407 nukleotidoko luzera) (Xiao et al., 2008a).

$\gamma 4$ familia behiaren ERV talderik ugariena da. Haren kopiak 1., 2., 6., 10., 15., 23., 26., 28. eta X kromosometan agertu ziren. $\gamma 4$ familiako ERV prototipoaren genomak 8.576 base-pare ditu eta eretrobirusen ohiko egitura du, *gag* (486 aminoazidoko luzera), *pro* (271 aa), *pol* (1416 aa) eta *env* (haren egitura ez dago argi) geneak eta, alboetan, LTRak (469 eta 470 nukleotidoko luzera) (Xiao et al., 2008c).

1.6.2. Equidae

Equidae familian, *Equus* generoa osatzen duten 8 espezie sailkatzen dira, hots, zaldiak, zebrak eta astoak. Animalia horiek kontinente gehienetan daude eta talde horren lehen fosilak orain dela 54 miloi urte agertu ziren. Familia horretako espezieak herbiboroak dira eta ekologia eta jokaera aldakorra azaltzen dituzte (Nowak, 1999).

1.6.2.1. Zaldia

Zaldia Equidae familia barruan sekuentziatu den lehen espezie da. Sekuentziazioa WGS metodoaren bidez egiten ari da eta, lan horretatik, EquCab1 eta EquCab2 genoma-bertsioak argitaratu dira; biek 6.8X-ko estaldura dute. Azken bertsioan, 31 kromosoma autosomikoren eta X kromosomaren informazioa eskuragarri jarri zen, guztira 2.3Gbp (Broad Institute, 2008).

1.6.2.2. ERVak Equidae-n

Equidae familian edota zaldietan ERVak gutxi aztertu dira. Talde horretan ERVei buruz ezaguna dena ERVei buruzko lan orokorretatik dakigu. Alde batetik, ungaratuetan II. klaseko ERVak detektatzeko lana dago (Hecht et al., 1996). Lan horretan, Southern blot bidez, zaldian hibridazio

ahul batzuk lortu ziren; hots, ERVak bazirela detektatu zen (Hecht et al., 1996). Bestetik, III. klaseko ERVak mamalioetan ezberdinetan detektatzeko lana dago (Benit et al., 1999). Zaldiaren kasuan, Southern blot teknika erabilia, hainbat hibridazio lortu ziren eta 5 ERVen *pol* genearen sekuentziak lortzeko gai izan ziren (Benit et al., 1999).

I.7. Lanaren planteamendua

Sarreran, kapitulu ezberdinetan zehar, egun erretrovirus endogenoi (ERV) buruz eta ERVekin lotutako hainbat alderdiri buruz ezaguna dena aztertu dugu. Hala ere, oraindik ERVei buruz asko argitzeke dago.

Zientzian, oro har, hipotesiek zuzenduta (deduktiboa) lan egiten bada ere, genomika osteko aroan emaitzek zuzenduta (induktiboa) lan egiten da. Horrela lan egiteak hainbat kritika jaso baditu ere, lan egiteko bi hurbilketak osagarriak direla defendatzen da (Kell & Oliver, 2004). Kontutan hartu behar da genomika osteko aroan behaketa asko egiteke daudela eta behaketa horiek egin ostean hipotesien planteamendu hasi daitekeela. Beraz, emaitzak hipotesiek zuzendu gabe sortzen dira behaketak lortu ahal izateko eta, horrela, hipotesiei bide eman (Kell & Oliver, 2004). Hortaz, hemen egingo den lanaren planteamendua, hipotesietan baino, emaitzak (behaketak) sortzeko abiapuntuetan oinarritzen da.

ERVen banaketa eta ERV klase bakoitzaren banaketa ornodunetan ezaguna bada ere (Herniou et al., 1998; Benit et al., 1999; Gifford & Tristem, 2003; Gifford et al., 2005), mamalioetan, genoma osoko bilaketak eta karakterizazioak primate (Tristem, 2000; Paces et al., 2002; Villesen et al., 2004; Polavarapu et al., 2006) eta karraskarietara mugatu da (Baillie et al., 2004; McCarthy & McDonald, 2004). Hau da, gainontzeko mamalioetan ERVen presentzia ezaguna bada ere, I klasekoena batez ere, ERV kopurua, ERV klase bakoitzaren presentzia edota ERVen banaketa genomak zehar ezezaguna da. Genoma osoko bilaketak egiteko lanabes bioinformatikoak eraginkorrak izan dira eta, mota honetako metodologiak erabili ahal izateko, sekuentziazio maila altua duten genomak beharrezkoak dira. Irizpide hau betetzen duten genoma sekuentziazio proiektuen artean behia (*Bos taurus*) eta zaldia (*Equus caballus*) aurkitzen dira (7.2X eta 6.8X alditan sekuentziatuak hurrenez hurren). Bestalde, bi animalia hauek ERVak bioinformatikoki aztertu ez diren bi taxoien kide dira, hots, Ruminantia eta Equidae taxoienak hurrenez hurren. Hala ere, Ruminantia taxoian behietan (Xiao et al., 2008b) eta ardietan (Klymiuk et al., 2003) lan esperimentalak egin dira. Azkenik, orain arte egindako genoma osoko lanetan, talde taxonomikoen artean ezberdintasunak aurkitu dira. Hortaz, aurretik ezaguna dena kontutan hartuz, “*Erretrovirus endogenoak behiaren eta zaldiaren genomaren parte dira, genoma bakoitzean kopuru aldakorrean agertzen dira eta erretrovirus endogenoen klase bakoitzak presentzia ezberdina izango badu ere, batez ere I klasekoak gailenduak dira. Gainera bi genomak artean ezberdintasunak daude.*” abiapuntutzat proposa daiteke.

ERVen beste alderdi interesgarri bat euren dinamika ebolutiboa da. Genomaren zati diren heinean, genomak jasotzen dituen presio ebolutiboen menpe gelditzen dira eta, askotan, funtzionaltasuna galtzeko joera dute. Hala ere, ERV ezberdinek funtzioa mantentzeko hautespena jasotzen dute (Gifford & Tristem, 2003; Jern & Coffin, 2008). Bestalde, ERV familia ezberdinek ostalariaren genomak euren kopurua mekanismo ezberdinen bidez kopia-kopurua emendatzeko gaitasuna azaldu dute (Belshaw et al., 2004; Belshaw et al., 2005; Katzourakis et al., 2005). Azkenik, ERV ezberdinek euren artean errekonbinatzeko gaitasuna azaldu dute (Klymiuk et al., 2006; Lee et al., 2008; Kjeldbjerg et al., 2008). Hala ere, dinamika ebolutiboa osatzen duten mekanismo ezberdin hauek soilik gizakietan eta, neurri txikiago batean, txerrietan, aztertutakoak izan dira. Beraz, behian zein zaldian detekta daitezkeen ERVetan euren dinamika ebolutiboa aztertzeak beste taxoi batzuetan gertatzen dena ezagutzea baimendu liguke. Gainera, beste taxoi batzuetan gertatzen denarekin erkatzeko, sekuentziazio maila altua duten beste mamalioetan ERVen dinamika aztertu behar da. Hortaz “*behiaren eta zaldiaren erretrovirus endogenoen dinamika ebolutiboa gainontzeko mamalioen ERVen dinamika ebolutiboak dituen antzeko ezaugarriak dituzte, hau da, antzeko hautespena, kopia-kopuruaren emendio-*

mekanismo eta errekonbinazioa azaltzen dute.” abiapuntutzat proposa daiteke.

ERVen txertaketa zorizkoa den ala tropismoren bat azaltzen duten eta txertaketa horrek ostalariaren genomak duen eragina eztabaidatua izan da (Gifford & Tristem, 2003; Jern & Coffin, 2008). Hainbat espezieetan elementu transposakorrak edota LTR erretrotransposoiak ostalariaren geneetan txertatu direla ikusi da, batez ere geneen introietan (Nekrutenko & Li, 2001; Medstrand et al., 2002; Ganko et al., 2003; van de Lagemaat et al., 2003; Jordan et al., 2003; DeBarry et al., 2006; Ganko et al., 2006; Almeida et al., 2007). Gainera, hainbat espezieetan gene hauek funtzio jakin batzuekin lotu dira Gene Ontology proiektuaren anotazioa jarraituz, adibidez, gizakian eta saguan estresaren aurreko erantzunarekin, kanpo-kinaden aurreko erantzunarekin edo defentsarekin (van de Lagemaat et al., 2003; DeBarry et al., 2006; Ganko et al., 2006). Hala ere alderdi honi buruz ERVetan gutxi da ezaguna. Gene Ontology-ren anotazio maila dela eta, azterketa hau soilik behian burutu daiteke. Hortaz *“behiaren geneetan edo geneetatik gertu behiaren erretrobirus endogenoak txertatuta aurkitzen dira”* abiapuntutzat proposa daiteke.

Azkenik, hainbat espezieetan ERVak adierazten direla ikusi da eta hainbat espezieetan elementu transposakorrak ere adierazten direla (Spencer et al., 2003; Stauffer et al., 2004; Okahara et al., 2004; Seifarth et al., 2005; Klymiuk et al., 2006). Are gehiago, gizakien ERVekin egindako lanetan, ERV familia ezberdinak, ehun eta garapen-fase ezberdinetan adierazten direla ikusi da (Stauffer et al., 2004; Okahara et al., 2004; Seifarth et al., 2005). Hala ere, ERVen adierazpen sistematikoa ez da gizakietan baino aztertu eta erabili diren metodoen artean, bionformatikaren bidez EST datubaseak erabiliz ERVen adierazpena aztertu da. Beraz, beste espezieen ERVen adierazpena aztertzea interesgarria izango litzateke ERVei buruzko ezagumenduan sakontzeko. Tamalez, ERVen adierazpena aztertzeko erabiltzen den EST datubasea, zaldi jatorria duten sekuentziaz ez dago oso hornitua, beraz, analisi hau behian detektatu daitezkeen ERVekin soilik burutu daiteke. Hortaz, *“behiaren erretrobirus endogenoak adierazten dira eta erretrobirus endogenoen familia bakoitza ehun eta garapen fase ezberdinetan modu ezberdinean adierazten da.”* abiapuntutzat proposa daiteke.

1.7.1. Helburu espezifikoak

Lan honen helburua behiaren (*Bos taurus*) eta zaldiaren (*Equus caballus*) genomak aztertuz eta lanabes bioinformatikoak erabiliz mamalioen erretrobirus endogenoei buruzko ezagutza sakontzea da. Horretarako:

- a) Behiaren eta zaldiaren genomak erretrobirus endogenoak identifikatuko eta karakterizatuko dira.
- b) Behiaren eta zaldiaren erretrobirus endogenoen dinamika ebolutiboa aztertuko da, erretrobirus endogenoen hautespena, kopia-kopuruaren emendio-mekanismoa eta errekonbinazioan sakonduz, eta gainontzeko mamalioekin erkatuko da.
- c) Behiaren erretrobirus endogenoetatik gertu dauden geneak aztertuko eta karakterizatuko dira.
- d) Behiaren erretrobirus endogenoen adierazpena detektatuko da ESTak erabiliz.

Lan hau kapitulutan antolatu da. Hortaz, (a) helburua “Genome-wide detection and characterization of endogenous retroviruses in *Bos taurus*” eta “Zaldiaren erretrobirus endogenoen (ERV) detekzioa eta sailkapena” ataletan (II eta III atalak hurrenez hurren) jorratuko da, (b) helburua “Mamalioen erretrobirus endogenoen (ERV) dinamika ebolutiboa” atalean (IV atala), (c) helburua “Genes around bovine endogenous retroviruses (BoERVs) are related to host defense and histone functions” atalean (V atala) eta (d) helburua “Behiaren erretrobirus endogenoen (BoERV) adierazpena behiaren (*Bos taurus*) EST-etan” atalean (VI atala).

I.8. Bibliografia

Almeida, L., I. Silva, W. Silva, J. Castro, P. Riggs, C. Carareto, and M. Amaral. 2007. The contribution of transposable elements to *Bos taurus* gene structure. *Gene* **390**:180-189.

Altschul, S., T. Madden, A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D. Lipman. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* **25**:3389-3402.

Andersson, A., P. Venables, R. Tonjes, J. Scherer, L. Eriksson, and E. Larsson. 2002. Developmental expression of HERV-R (ERV3) and HERV-K in human tissue. *Virology* **297**:220-225.

Arnaud, F., M. Varela, T. Spencer, and M. Palmarini. 2008. Coevolution of endogenous Betaretroviruses of sheep and their host. *Cellular and Molecular Life Sciences* **65**:3422-3432.

Baillie, G., L. van de Lagemaat, C. Baust, and D. Mager. 2004. Multiple groups of endogenous betaretroviruses in mice, rats, and other mammals. *Journal of Virology* **78**:5784-5798.

Bannert, N. and R. Kurth. 2004. Retroelements and the human genome: New perspectives on an old relation. *Proceedings of the National Academy of Sciences of the United States of America* **101**:14572-14579.

Belshaw, R., A. Katzourakis, J. Paces, A. Burt, and M. Tristem. 2005. High copy number in human endogenous retrovirus families is associated with copying mechanisms in addition to reinfection. *Mol Biol Evol* **22**:814-817.

Belshaw, R., V. Pereira, A. Katzourakis, G. Talbot, J. Paces, A. Burt, and M. Tristem. 2004. Long-term reinfection of the human genome by endogenous retroviruses. *Proceedings of the National Academy of Sciences of the United States of America* **101**:4894-4899.

Benit, L., P. Dessen, and T. Heidmann. 2001. Identification, phylogeny, and evolution of retroviral elements based on their envelope genes. *Journal of Virology* **75**:11709-11719.

Benit, L., J. Lallemand, J. Casella, H. Philippe, and T. Heidmann. 1999. ERV-L elements: a family of endogenous retrovirus-like elements active throughout the evolution of mammals. *Journal of Virology* **73**:3301-3308.

Biemont, C. and C. Vieira. 2006. Genetics: Junk DNA as an evolutionary force. *Nature* **443**:521-524.

Blikstad, V., F. Benachenhou, G. Sperber, and J. Blomberg. 2008. Evolution of human endogenous retroviral sequences: a conceptual account. *Cellular and Molecular Life Sciences* **65**:3348-3365.

Blomberg, J., F. Benachenhou, V. Blikstad, G. r. Sperber, and J. Mayer. 2009. Classification and nomenclature of endogenous retroviral sequences (ERVs): Problems and recommendations. *Gene*:448:115-123.

Boeke, J. and J. Stoye. 1997. Retrotransposons, endogenous retroviruses, and the evolution of retroelements, p. 343-435. *In* J. Coffin, S. Hughes, and H. Varmos (eds.), *Retroviruses*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Broad Institute. Horse Genome Project. <http://www.broadinstitute.org/mammals/horse> . 2008.

Ref Type: Electronic Citation

Bushman, F., M. Lewinski, A. Ciuffi, S. Barr, J. Leipzig, S. Hannenhalli, and C. Hoffmann. 2005. Genome wide analysis of retroviral DNA integration. *Nature Reviews Microbiology* **3**:848-858.

Costas, J. 2001. Evolutionary dynamics of the human endogenous retrovirus family HERV-K inferred from full-length proviral genomes. *Journal of Molecular Evolution* **53**:237-243.

Dangel, A., B. Baker, A. Memndoza, and C. Yu. 1995. Complement component C4 gene intron-9 as a phylogenetic marker for primates and indicator of the great ape slow-down phenomenon - Long, Terminal Repeats of the endogenous retrovirus ERV-K(C4) are a molecular clock of evolution. *American Journal of Human Genetics* **57**:921.

DeBarry, J., E. Ganko, E. McCarthy, and J. McDonald. 2006. The contribution of LTR retrotransposon sequences to gene evolution in *Mus musculus*. *Mol Biol Evol* **23**:479-481.

Denesvre, C., D. Soubieux, G. Pin, D. Hue, and G. Dambrine. 2003. Interference between avian endogenous ev/J 4.1 and exogenous ALV-J retroviral envelopes. *J Gen Virol* **84**:3233-3238.

Duret, L., G. Marais, and C. Biemont. 2000. Transposons but Not Retrotransposons Are Located Preferentially in Regions of High Recombination Rate in *Caenorhabditis elegans*. *Genetics* **156**:1661-1669.

Eickbush, T. H. and A. V. Furano. 2002. Fruit flies and humans respond differently to retrotransposons. *Current Opinion in Genetics & Development* **12**:669-674.

Ellinghaus, D., S. Kurtz, and U. Willhoeft. 2008. LTRharvest, an efficient and flexible software for de novo detection of LTR retrotransposons. *Bmc Bioinformatics* **9**.

Elsik, C., R. Tellam, K. Worley, and Bovine Genome Sequencing Anal Cons. 2009. The Genome Sequence of Taurine Cattle: A Window to Ruminant Biology and Evolution. *Science* **324**:522-528.

Fernandez, M. and E. Vrba. 2005. A complete estimate of the phylogenetic relationships in Ruminantia: a dated species-level supertree of the extant ruminants. *Biological Reviews* **80**:269-302.

Frank, O., M. Giehl, C. Zheng, R. Hehlmann, C. Leib-Mosch, and W. Seifarth. 2005. Human endogenous retrovirus expression profiles in samples from brains of patients with schizophrenia and bipolar disorders. *Journal of Virology* **79**:10890-10901.

Frank, O., C. Verbeke, N. Schwarz, J. Mayer, A. Fabarius, R. Hehlmann, C. Leib-Mosch, and W. Selfarth. 2008. Variable transcriptional activity of endogenous retroviruses in human breast cancer. *Journal of Virology* **82**:1808-

1818.

Ganko, E., V. Bhattachajee, P. Schliekelman, and J. McDonald. 2003. Evidence for the contribution of LTR retrotransposons to C-elegans gene evolution. *Mol Biol Evol* **20**:1925-1931.

Ganko, E., C. Greene, J. Lewis, V. Bhattacharjee, and J. McDonald. 2006. LTR retrotransposon-gene associations in *Drosophila melanogaster*. *Journal of Molecular Evolution* **62**:111-120.

Gifford, R., P. Kabat, J. Martin, C. Lynch, and M. Tristem. 2005. Evolution and distribution of class II-related endogenous retroviruses. *Journal of Virology* **79**:6478-6486.

Gifford, R. and M. Tristem. 2003. The evolution, distribution and diversity of endogenous retroviruses. *Virus Genes* **26**:291-316.

Gifford, R., A. Katzourakis, M. Tristem, O. Pybus, M. Winters, and R. Shafer. 2008. A transitional endogenous lentivirus from the genome of a basal primate and implications for lentivirus evolution. *Proceedings of the National Academy of Sciences of the United States of America* **105**:20362-20367.

Goff, S. 2004. Retrovirus restriction factors. *Molecular Cell* **16**:849-859.

Golan, M., A. Hizi, J. Resau, N. Yaal-Hahoshen, H. Reichman, I. Keydar, and I. Tsarfaty. 2008. Human endogenous retrovirus (HERV-K) reverse transcriptase as a breast cancer prognostic marker. *Neoplasia* **10**:521-5U3.

Griffiths, D. J. 2001. Endogenous retroviruses in the human genome sequence. *Genome Biol* **2**:REVIEWS1017.

Han, K., M. Konkel, J. Xing, H. Wang, J. Lee, T. Meyer, C. Huang, E. Sandifer, K. Hebert, E. Barnes, R. Hubley, W. Miller, A. Smit, B. Ullmer, and M. Batzer. 2007. Mobile DNA in old world monkeys: A glimpse through the rhesus macaque genome. *Science* **316**:238-240.

Hecht, S., K. Stedman, J. Carlson, and J. DeMartini. 1996. Distribution of endogenous type B and type D sheep retrovirus sequences in ungulates and other mammals. *Proceedings of the National Academy of Sciences of the United States of America* **93**:3297-3302.

Herbst, H., M. Sauter, and N. MuellerLantzsch. 1996. Expression of human endogenous retrovirus K elements in germ cell and trophoblastic tumors. *American Journal of Pathology* **149**:1727-1735.

Herniou, E., J. Martin, K. Miller, J. Cook, M. Wilkinson, and M. Tristem. 1998. Retroviral diversity and distribution in vertebrates. *Journal of Virology* **72**:5955-5966.

Huda, A., N. Polavarapu, I. Jordan, and J. McDonald. 2008. Endogenous retroviruses of the chicken genome. *Biology Direct* **3**.

Hughes, J. and J. Coffin. 2005. Human endogenous retroviral elements as indicators of ectopic recombination events in the primate genome. *Genetics* **171**:1183-1194.

International Human Genome Sequencing Consortium. 2001. Initial sequencing and analysis of the human genome. *Nature* **409**:860-921.

Jern, P. and J. Coffin. 2008. Effects of Retroviruses on Host Genome Function. *Annual Review of Genetics* **42**:709-732.

Johnson, W. and J. Coffin. 1999. Constructing primate phylogenies from ancient retrovirus sequences. *Proceedings of the National Academy of Sciences of the United States of America* **96**:10254-10260.

Jordan, I. K., I. B. Rogozin, G. V. Glazko, and E. V. Koonin. 2003. Origin of a substantial fraction of human regulatory sequences from transposable elements. *Trends in Genetics* **19**:68-72.

Jurka, J. 2000. Repbase Update - a database and an electronic journal of repetitive elements. *Trends in Genetics* **16**:418-420.

Kalyanaraman, A. and S. Aluru. 2005. Efficient algorithm's and software for detection of full-length LTR retrotransposons. 2005 IEEE Computational Systems Bioinformatics Conference, Proceedings56-64.

Katzourakis, A., V. Pereira, and M. Tristem. 2007a. Effects of recombination rate on human endogenous retrovirus fixation and persistence. *Journal of Virology* **81**:10712-10717.

Katzourakis, A., A. Rambaut, and O. Pybus. 2005. The evolutionary dynamics of endogenous retroviruses. *Trends in Microbiology* **13**:463-468.

Katzourakis, A., M. Tristem, O. Pybus, and R. Gifford. 2007b. Discovery and analysis of the first endogenous lentivirus. *Proceedings of the National Academy of Sciences of the United States of America* **104**:6261-6265.

Kell, D. and S. Oliver. 2004. Here is the evidence, now what is the hypothesis? The complementary roles of inductive and hypothesis-driven science in the post-genomic era. *Bioessays* **26**:99-105.

Kjeldbjerg, A., P. Villesen, L. Aagaard, and F. Pedersen. 2008. Gene conversion and purifying selection of a placenta-specific ERV-V envelope gene during simian evolution. *Bmc Evolutionary Biology* **8**.

Klymiuk, N., M. Muller, G. Brem, and B. Aigner. 2003. Characterization of endogenous retroviruses in sheep. *Journal of Virology* **77**:11268-11273.

Klymiuk, N., M. Muller, G. Brem, and B. Aigner. 2006. Phylogeny, recombination and expression of porcine endogenous retrovirus {gamma}2 nucleotide sequences. *J Gen Virol* **87**:977-986.

Lee, D., J. Lee, N. Park, Y. Oh, M. Kwon, and Y. Kim. 2008. Analysis of natural recombination in porcine endogenous retrovirus envelope genes. *Journal of Microbiology and Biotechnology* **18**:585-590.

Madden, T. 2002. The BLAST Sequence Analysis Tool *In* J. McEntyre and J. Ostell (eds.), *The NCBI Handbook*.

National Library of Medicine (US), NCBI, Bethesda (MD).

Maksakova, I., D. Mager, and D. Reiss. 2008. Keeping active endogenous retroviral-like elements in check: the epigenetic perspective. *Cellular and Molecular Life Sciences* **65**:3329-3347.

Martin, J., E. Herniou, J. Cook, R. O'Neill, and M. Tristem. 1999. Interclass transmission and phyletic host tracking in murine leukemia virus-related retroviruses. *Journal of Virology* **73**:2442-2449.

McCarthy, E. and J. McDonald. 2003. LTR_STRUC: a novel search and identification program for LTR retrotransposons. *Bioinformatics* **19**:362-367.

McCarthy, E. and J. McDonald. 2004. Long terminal repeat retrotransposons of *Mus musculus*. *Genome Biology* **5**.

McClure, M., H. Richardson, R. Clinton, C. Hepp, B. Crowther, and E. Donaldson. 2005. Automated characterization of potentially active retroid agents in the human genome. *Genomics* **85**:512-523.

Medstrand, P., L. van de Lagemaat, and D. Mager. 2002. Retroelement distributions in the human genome: Variations associated with age and proximity to genes. *Genome Research* **12**:1483-1495.

Mouse Genome Sequencing Consortium. 2002. Initial sequencing and comparative analysis of the mouse genome. *Nature* **420**:520-562.

Nekrutenko, A. and W. H. Li. 2001. Transposable elements are found in a large number of human protein-coding genes. *Trends in Genetics* **17**:619-621.

Nowak, R. 1999. WALKERS MAMMALS WORLD 2. The Johns Hopkins University Press, Baltimore.

Okahara, G., S. Matsubara, T. Oda, J. Sugimoto, Y. Jinno, and F. Kanaya. 2004. Expression analyses of human endogenous retroviruses (HERVs): tissue-specific and developmental stage-dependent expression of HERVs. *Genomics* **84**:982-990.

Paces, J., A. Pavlicek, and V. Paces. 2002. HERVd: database of human endogenous retroviruses. *Nucleic Acids Research* **30**:205-206.

Perron, H., C. Bernard, J. B. Bertrand, A. B. Lang, I. Popa, K. Sanhadji, and J. Portoukalian. 2009. Endogenous retroviral genes, Herpesviruses and gender in Multiple Sclerosis. *Journal of the Neurological Sciences* **286**:65-72.

Polavarapu, N., N. Bowen, and J. McDonald. 2006. Identification, characterization and comparative genomics of chimpanzee endogenous retroviruses. *Genome Biology* **7**.

Prusty, B., H. zur Hausen, R. Schmidt, R. Kimmel, and E. de Villiers. 2008. Transcription of HERV-E and HERV-E-related sequences in malignant and non-malignant human haematopoietic cells. *Virology* **382**:37-45.

Rat Genome Sequencing Project Cons. 2004. Genome sequence of the Brown Norway rat yields insights into mammalian evolution. *Nature* **428**:493-521.

Samuelson, L., K. Wiebauer, C. Snow, and M. Meisler. 1990. Retroviral and Pseudogene Insertion Sites Reveal the Lineage of Human Salivary and Pancreatic Amylase Genes from A Single Gene During Primate Evolution. *Molecular and Cellular Biology* **10**:2513-2520.

Seifarth, W., O. Frank, U. Zeilfelder, B. Spiess, A. Greenwood, R. Hehlmann, and C. Leib-Mosch. 2005. Comprehensive analysis of human endogenous retrovirus transcriptional activity in human tissues with a retrovirus-specific microarray. *Journal of Virology* **79**:341-352.

Smit, AFA, Hubley, R, and Green, P. RepeatMasker Open-3.0. 2004.
Ref Type: Internet Communication

Spencer, T., M. Mura, C. Gray, P. Griebel, and M. Palmarini. 2003. Receptor usage and fetal expression of ovine endogenous betaretroviruses: Implications for coevolution of endogenous and exogenous retroviruses. *Journal of Virology* **77**:749-753.

Sperber, G., T. Airola, P. Jern, and J. Blomberg. 2007. Automated recognition of retroviral sequences in genomic data - RetroTector (c). *Nucleic Acids Research* **35**:4964-4976.

Sperber, G., A. Lovgren, N. E. Eriksson, F. Benachenhou, and J. Blomberg. 2009. RetroTector online, a rational tool for analysis of retroviral elements in small and medium size vertebrate genomic sequences. *Bmc Bioinformatics* **10**:S4.

Stauffer, Y., G. Theiler, P. Sperisen, Y. Lebedev, and C. V. Jongeneel. 2004. Digital expression profiles of human endogenous retroviral families in normal and cancerous tissues. *Cancer Immun* **4**:2.

Stengel, A., C. Roos, G. Hunsmann, W. Seifarth, C. Leib-Mosch, and A. Greenwood. 2006. Expression profiles of endogenous retroviruses in old world monkeys. *Journal of Virology* **80**:4415-4421.

Tarlinton, R., J. Meers, and P. Young. 2008. Biology and evolution of the endogenous koala retrovirus. *Cellular and Molecular Life Sciences* **65**:3413-3421.

Tarlinton, R., J. Meers, and P. Young. 2006. Retroviral invasion of the koala genome. *Nature* **442**:79-81.

Tristem, M. 2000. Identification and characterization of novel human endogenous retrovirus families by phylogenetic screening of the Human Genome Mapping Project database. *Journal of Virology* **74**:3715-3730.

van de Lagemaat, L., J. Landry, D. Mager, and P. Medstrand. 2003. Transposable elements in mammals promote regulatory variation and diversification of genes with specialized functions. *Trends in Genetics* **19**:530-536.

- van de Lagemaat, L., P. Medstrand, and D. Mager.** 2006. Multiple effects govern endogenous retrovirus survival patterns in human gene introns. *Genome Biology* **7**.
- Villesen, P., L. Aagaard, C. Wiuf, and F. S. Pedersen.** 2004. Identification of endogenous retroviral reading frames in the human genome. *Retrovirology* **1**.
- Vogt, V.** 1997. Retroviral virions and genomes, p. 343-435. *In* J. Coffin, S. Hughes, and H. Varmos (eds.), *Retroviruses*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Wilson, C.** 2008. Porcine endogenous retroviruses and xenotransplantation. *Cellular and Molecular Life Sciences* **65**:3399-3412.
- Xiao, R., J. Kim, H. Choi, K. Park, H. Lee, and C. Park.** 2008a. Characterization of the bovine endogenous retrovirus beta 3 genome. *Molecules and Cells* **25**:142-147.
- Xiao, R., K. Park, H. Lee, J. Kim, and C. Park.** 2008b. Identification and classification of endogenous retroviruses in cattle. *Journal of Virology* **82**:582-587.
- Xiao, R., K. Park, Y. Oh, J. Kim, and C. Park.** 2008c. Structural Characterization of the Genome of BERV gamma 4, the Most Abundant Endogenous Retrovirus Family in Cattle. *Molecules and Cells* **26**:404-408.
- Yoder, J., C. Walsh, and T. Bestor.** 1997. Cytosine methylation and the ecology of intragenomic parasites. *Trends in Genetics* **13**:335-340.

II. Genome-wide detection and characterization of endogenous retroviruses in *Bos taurus*

II.1. Abstract

Endogenous retroviruses (ERVs) are the proviral phase of exogenous retroviruses that were integrated into the host germ line. ERVs have been detected in wide range of vertebrates and they can play an important role in their host genome. In order to detect ERVs genome-wide analyses have been performed using bioinformatic tools in primates and rodents. However, the knowledge of ERVs in other mammal groups is not so wide and the source of information is basically experimental. Thus, in this work, cattle genome was analyzed with three different bioinformatic methods to improve the knowledge about ERVs. Each method used had a different efficiency detecting ERVs: BLAST-based detected 928 possible ERVs, LTR_STRUC 4487 ERVs and Retrorector© 9698 ERVs. So for new ERV detection the use of different methods is advisable due the differences in performance of each method. The detected ERVs did not show a homogeneous chromosomal distribution and the number of ERVs was positively correlated with chromosomal size and negatively with GC content of chromosome. Some of these ERVs were classified in 24 putative ERV families, 20 of them not described before. One of these new families, BoERV1, was the most abundant family and it seemed to be a ruminant specific ERV family. Thus, the number of ERV families in cow is not as limited as it was previously thought.

Finally, the representative of ERV families from rodents, primates and ruminants showed a phylogenetic relationship following their hosts' relationship.

II.2. Introduction

Endogenous retroviruses (ERVs) are the proviral phase of exogenous retroviruses that were integrated into the host germ line and remained integrated into the host genome for generations. ERVs have been detected in all mammals and a wide range of other vertebrates and their typical structure comprises a central part with the three major genes (*gag*, *pol* and *env*) flanked by 2 LTRs (Long Terminal Repeats) which were identical at the time of integration (Boeke & Stoye, 1997).

The biological significance of retrotransposons ranges from their contribution to mutation and disease to their role in gene and genome evolution. For example, in humans, mice and sheep an *env* gene of retroviral origin, acquired independently in the different species, is involved in placenta morphogenesis. (Jern & Coffin, 2008).

ERV-mediated recombination events have had profound effects in the shaping of the host's genome, and new ERV integrations introduce added variation to the host transcriptomes (Jern & Coffin, 2008).

The expression of ERVs has been associated with several positive physiological functions as well as certain diseases. Their roles as aetiological agents, possible contributing factors or markers of disease have been well established in experimental animals (Jern & Coffin, 2008), and in recent years more and more in humans. For example, the expression of ERVs in humans has been connected with poor prognosis in breast cancer (Golan et al., 2008), it has been discovered that a human ERV is involved in melanoma cell malignant transformation (Serafino et al., 2009) and it has been suggested that a human ERV may have a role in multiple sclerosis (Perron et al., 2009).

At present, there is no well-established or accepted standard for naming and classifying all ERVs. For human ERVs (HERVs) the tRNA complementary to the PBS has traditionally been used for this purpose. This classification is, however, unreliable as proviruses from the same phylogenetic groups may display differences in PBS and otherwise unrelated proviruses may use the same tRNA as a primer. In fact, phylogenetic and related comparisons have proved to be more useful for classification of ERVs (Jern & Coffin, 2008). So ERVs are classified in three classes depending on the phylogenetic relationship with the exogenous retrovirus genus: Class I ERVs are related to the *Gammaretrovirus* and *Epsilonretrovirus* genera; Class II are related to the *Alfaretrovirus*, *Betaretrovirus*, *Deltaretrovirus* and *Epsilonretrovirus* genera; and Class III is related to the *Spumavirus* genus (Gifford et

al., 2005).

Since whole-genome sequences became available, genome-wide analyses for detection of ERVs have been performed using bioinformatic tools. The first studies in humans were conducted using BLAST-based searches (Tristem, 2000; Villesen et al., 2004), and the same method was used in rodents (Baillie et al., 2004). For the study of the chimpanzee genome a new program called LTR_STRUC (McCarthy & McDonald, 2003) was applied, combined with BLAST-based search (Polavarapu et al., 2006). Recently, a specific new program for ERV detection, Retrotector©, has been published (Sperber et al., 2007). In the last years the first overview of ERVs in whole genome sequences has been generated by RepeatMasker (Smit et al., 2004), a program designed to identify repetitive elements using the RepBase database (Jurka, 2000).

The mammalian order Cetartiodactyla has become a major focus of attention in comparative genomics, because it comprises a phylogenetically distant clade of eutherian mammals related to primates which diverged from a common ancestor ~85 million years ago (Kumar & Hedges, 1998). *Bos taurus* is one of the world's most important food animal species and is also among the most biologically interesting due to its unique digestive, reproductive and immune system physiology. The unveiling of the cattle genome sequence in 2009 allowed the first comprehensive effort to be made to catalogue the diversity of transposable elements in the cattle genome (Elsik et al., 2009). Interspersed repeats cover 45.54% of the genome. Among them, non-LTR retrotransposons LINEs account for 23.39% of the genome and SINEs 17.66%. LTR retrotransposons, which include ERVs, account for 3.19% of the genome (Elsik et al., 2009).

In search of ERV elements, the cattle genome has also been analyzed experimentally. Using a PCR-based approach (Xiao et al., 2008b) a number of ERVs were detected and classified in 4 families named $\beta 3$, $\gamma 4$, $\gamma 7$ and $\gamma 9$, on the basis of their similarity to ovine ERVs (OERV). The structure and sequences of BERV- $\beta 3$ and the abundant BERV- $\gamma 4$ ERVs were also analysed. Based on these studies it was suggested that the ERV family expansion in cattle was more limited than in other artiodactyls such as pigs and sheep (Xiao et al., 2008a; Xiao et al., 2008b; Xiao et al., 2008c).

In this paper a process to detect ERVs in the cow genome is carried out. To that end three different methods were used: BLAST-based searches using retroviral sequences and the LTR_STRUC and Retrotector© programs. Those ERVs which were detected by at least 2 methods and whose RT region was longer than 500 nucleotides were used to define bovine ERV families. Finally, the representative ERV for each family and families from other species were used to study the relationship between the ERVs of different species.

II.3. Materials and methods

II.3.1. Genomic sequence

In this work the Btau_3.1 version of cow (*Bos Taurus*) genome (x7,1 cover) was analysed. It was retrieved from Baylor College of Medicine (<http://www.hgsc.bcm.tmc.edu/projects/bovine/index.html>).

II.3.2. Detection of BoERVs

Three different approaches were applied and compared.

1) The first strategy was based on the similarity of sequences. Segments of aminoacidic sequences from *gag* and *env* (TM region) genes from well-annotated 12 exogenous retroviruses were used as a search query (see Table II-1). In the case of *pol* gene (RT region) one piece for each retroviral class was used (from MLV, MPMV (Baillie et al., 2004) and BFV retroviruses). Individual gene segments were searched in each chromosome using tBLASTn implemented in NCBI-BLAST 2.2.14 program (Altschul et al., 1997). The results of this search were parsed by home-made scripts written in PHP.

To build the possible ERV elements the results of each gene were compared. Based on the 12 search-retrovirus the distances between gene queries were calculated (Table II-1. If, at least, two matches were in the limits of these distances the region was defined as a possible ERV.

2) In the second approach the program LTR_STRUC (McCarthy & McDonald, 2003) was used to find LTR elements. To find this elements this program searches for LTR pairs and applies supplementary analysis to confirm the element.

3) In the third strategy the program Retroector© (Sperber et al., 2007) was used to find possible ERV elements. Briefly, this program searches retroviral elements using motifs of different retroviral regions.

Table II-1: Previously characterized retroviruses from a variety of species used as query.

Name	Genus	Host	Accession Number (GenBank)	Query sequence position			Distances between query sequences			
				<i>gag</i>	<i>pol</i>	<i>env</i>	<i>start-gag</i>	<i>gag-pol</i>	<i>pol-env</i>	<i>env-end</i>
Class I										
GALV	<i>Gammaretrovirus</i>	<i>Hylobates</i>	NC_001885	1255-1888		6998-7555	1255	846	3508	533
MLV	<i>Gammaretrovirus</i>	Muridae	NC_001501	1020-1653	2520-3276	6920-7457	1020	867	3644	875
FeLV	<i>Gammaretrovirus</i>	Felidae	NC_001940	1516-2149		7323-7860	1516	822	3593	588
Class II										
JSRV	<i>Betaretrovirus</i>	<i>Ovis aries</i>	NC_001494	1347-1667		6502-7111	1347	1444	2830	351
MMTV	<i>Betaretrovirus</i>	Muridae	NC_001503	1181-1774		6786-7362	1181	1194	3077	1443
BLV	<i>Deltaretrovirus</i>	<i>Bos taurus</i>	NC_001414	790-1362		5760-6158	790	572	2874	2261
HTLV	<i>Deltaretrovirus</i>	<i>Homo sapiens</i>	NC_001436	889-1455		5765-6269	889	780	2786	2238
EIAV	<i>Lentivirus</i>	<i>Equus caballus</i>	NC_001450	945-1562		7235-7867	945	807	4107	492
HIV	<i>Lentivirus</i>	<i>Homo sapiens</i>	NC_001802	1035-1401		7727-8261	1035	765	4805	920
Visna	<i>Lentivirus</i>	<i>Ovis aries</i>	NC_001452	973-1566		8176-8806	973	620	5237	396
Class III										
HSRV	<i>Spumavirus</i>	<i>Homo sapiens</i>	NC_001795	2111-2750		8788-9399	2111	868	4456	2555
BFV	<i>Spumavirus</i>	<i>Bos taurus</i>	NC_001831	2016-2652	3487-4204	8903-9515	2016	835	4699	2487
Mean ± standard deviation							1373±564	868±217	3851±799	1410±962

II.3.3. Distribution of BoERVs

The homogeneity of the detected elements' distribution was tested chromosome by chromosome as proposed by Villesen et al (Villesen et al., 2004). The expected element number (based in the mean density and the length of a chromosome) was compared with the observed number by means of two statistics a χ^2 test and a G-test (Sokal & Rohls, 1969), both test with one degree of freedom.

$$\chi^2 = \sum [(Observed - expected)^2 / expected]$$

$$G = 2 \sum Observed * \ln(observed/expected)$$

The correlation analyses between the number of detected elements by each method and chromosome length, GC content and gene density were performed using R (R Development Core Team, 2008).

II.3.4. Classification of BoERVs

Relationship of the detected elements was determined by a phylogenetical analysis of the RT region. The 247 sequences with high degree of similarity in BLAST searches (more than 500 nucleotides) were used and the 12 retrovirus used in the search, detected 8 cow's ERV (Xiao et al., 2008b), 12 sheep's ERV (Klymiuk et al., 2003). Others retrovirus used in previous works, such as, BoEV (X99924), HERV-E (M10976), PERV (AJ293656), MPMV (NC_001550), IAPM (M17551), OMVV (NC_001511), FeFV (U78765), MuERV-L (Y12713), and ZAM (AJ000387) were used as outgroup. The sequences were aligned using MAFFT 5.861 program (Katoh et al., 2005) (FFT-NS-1 option) and cleaned with Gblocks 0.91 (Castresana, 2000) (minimum length of block 5, allowed gap position with half, minimum number of sequences for a flank position 146, maximum number of contiguous nonconserved positions 10). Three trees were constructed (a) using Neighbor-Joining method implemented in MEGA 3.1 (Kumar et al., 2004) (p-distance, pairwise deletion and 1000 bootstrap), (b) using maximum likelihood method implemented in Phyml 2.4.4 (Guindon & Gascuel, 2003) (1000 bootstrap, GTR+G model, $\alpha=2.71$, estimated by Modeltest 3.7 (Posada & Crandall, 1998)) and (c) using Bayesian inference method implemented in MrBayes 3.1 (Ronquist &

Huelsenbeck, 2003) (10^6 generations, GTR+G+I model).

The families were defined based on the support of phylogenetical trees. Recently it was proposed a nomenclature to standardize ERV's names (Blomberg et al., 2009). However we decided to name the families according to the previous nomenclature. So we named bovine endogenous retrovirus the ERVs described in this work, and the acronym used was BoERV (Bovine Endogenous Retrovirus) in order to avoid confusion with previous names (Xiao et al., 2008b).

To elucidate the insertion time of elements classified in families, divergence of LTRs estimated by LTR_STRUC and Retrotector© were used. The insertion time of each element was estimated applying a substitution-rate of $2.3 - 5.0 \times 10^{-9}$ (Johnson & Coffin, 1999) to the divergence.

The representative members of a family were chosen as the closest element to the consensus sequence of the family or, when the consensus was not possible, the element with less stop codons. To build the consensus sequence the aminoacidic sequences of the members of each family were aligned using ClustalW (Thompson et al., 1994) and the consensus was get by EMBOSS suite's cons program (Rice et al., 2000). The distances between individual sequences and consensus were calculates using MEGA 3.1 (No of differences and pairwise deletion options).

II.3.5. PCR amplification of BoERV1 family elements in sheep

In order to amplify bovine BoERV1 in sheep, 6 samples of sheep were used A PCR was performed with 5'-TGTGCTGAGACAGAGGAAGC-3' (forward) and 5'-CCTATGGCCCTAGTCCCTTC-3' (reverse) primers. PCR conditions consisted of 5 min at 94°C followed by 25 cycles of a 55°C annealing step for 30 s, polymerization at 72°C for 30 s, and denaturation at 94°C for 30 s and finally one cycle at 72°C for 7 min. Reaction conditions were as follows: 13,1 µL of water, 2 µL of buffer, 0.8 µL of MgCl₂, 0.3 µL of dNTP, 0.3 µL of each primer, 0.2µL of Taq and 3 µL of DNA.

II.3.6. Interspecies comparison

The relationships of the BoERV groups with other species' ERV groups were analyzed using a phylogenetic tree. We used 16 representatives sequences from human ERV's groups (Tristem, 2000), 38 from chimpanzee (Polavarapu et al., 2006), 20 and 7 from mouse (Baillie et al., 2004; McCarthy & McDonald, 2004), 7 from rat (Baillie et al., 2004), 14 from sheep (Klymiuk et al., 2003), 8 from pig (Patience et al., 2001) and 24 from cow (our data). The sequences were aligned using MAFFT (linsi option) and positions with gaps in more than half of sequences were eliminated. The tree was constructed by Bayesian inference implemented in MrBayes 3.1 (10^6 generations, RtREV matrix +G+I).

To elucidate the insertion time of human HERV-S71 element and chimp CERV3 element we used the LTR divergence as in Johnson et al. (1999) (Johnson & Coffin, 1999). The LTRs of each element were aligned with ClustalW and the distance was calculated using MEGA 3.1 (K2P substitution-model). Applying a substitution-rate of $2.3 - 5.0 \times 10^{-9}$ (Johnson & Coffin, 1999) the insertion time was estimated.

II.4. Results

II.4.1. Detection and distribution of ERVs in the cow genome

ERVs were detected by three different methods in the cattle genome. The number of possible ERVs detected was very different with each method: a BLAST-based search detected 928 ERVs, while LTR_STRUC 4487 elements flanked with LTRs and Retrotector© detected up to 9698 possible ERVs, 10 times more than the first method (Table II-2). Accordingly, only 172 elements were detected by all three methods (Figure II-1). 739 of the elements detected by the BLAST-based search were also detected by Retrotector©. This program detected more elements than

LTR_STRUC and Retrorector© together (604 elements). Most of the elements (8183 elements) were detected by Retrorector© only (Figure II-1).

Table II-2: ERV elements detected by each method.

Chr.	Analyzed length (Mb)	BLAST				LTR_STRUC				Retrorector©			
		n	den	χ^2 test p-value ^a	G-test p-value	n	den	χ^2 test p-value ^a	G-test p-value	n	den	χ^2 test p-value ^a	G-test p-value
1	147	58	0,39	0,7533	0,7548	310	2,11	0,0108*	0,0126*	727	4,95	6,3E-10***	2,44E-09***
2	126	43	0,34	0,4790	0,4718	263	2,09	0,0303*	0,0339*	554	4,40	0,0117*	0,0131*
3	117	50	0,43	0,3849	0,3940	250	2,14	0,0128*	0,0152*	540	4,62	0,0002***	0,0003***
4	111	46	0,41	0,5361	0,5419	201	1,81	0,8601	0,8599	458	4,13	0,3727	0,3758
5	119	42	0,35	0,6349	0,6310	214	1,80	0,7749	0,7743	497	4,18	0,2273	0,2312
6	112	49	0,44	0,3040	0,3152	231	2,06	0,0661	0,0714	532	4,75	1,8E-05***	3,09E-05***
7	101	35	0,35	0,5874	0,5820	193	1,91	0,5544	0,5571	406	4,02	0,7640	0,7646
8	104	33	0,32	0,2957	0,2824	199	1,91	0,5352	0,5380	428	4,12	0,4206	0,4235
9	96	34	0,35	0,6858	0,6825	201	2,09	0,0541	0,0595	424	4,42	0,0222*	0,0247*
10	96	30	0,31	0,2797	0,2653	194	2,02	0,1653	0,1720	370	3,85	0,5891	0,5878
11	102	32	0,31	0,2734	0,2594	187	1,83	0,9975	0,9975	375	3,68	0,1394	0,1348
12	78	37	0,47	0,1648	0,1809	164	2,10	0,0738	0,0805	334	4,28	0,1484	0,1536
13	84	22	0,26	0,0759	0,0604	133	1,58	0,0855	0,0785	283	3,37	0,0054**	0,0044**
14	83	17	0,20	0,0087**	0,0041**	152	1,83	0,9912	0,9912	324	3,90	0,7871	0,7866
15	76	24	0,32	0,3626	0,3488	154	2,03	0,2058	0,2133	326	4,29	0,1445	0,1498
16	73	25	0,34	0,6059	0,6000	116	1,59	0,1181	0,1099	271	3,71	0,2773	0,2723
17	71	23	0,32	0,4438	0,4324	108	1,52	0,0489*	0,0425*	268	3,77	0,4218	0,4182
18	63	36	0,57	0,0120	0,0192*	91	1,44	0,0210*	0,0166*	191	3,03	0,0001***	9,00E-05***
19	64	27	0,42	0,5732	0,5798	83	1,30	0,0013**	0,0007***	174	2,72	4,1E-07***	8,28E-08***
20	69	16	0,23	0,0439*	0,0301*	126	1,83	0,9660	0,9660	254	3,68	0,2351	0,2296
21	64	25	0,39	0,8792	0,8798	107	1,67	0,3348	0,3277	240	3,75	0,3887	0,3845
22	60	15	0,25	0,1001	0,0798	70	1,17	0,0001***	3,60E-05***	171	2,85	1,2E-05***	4,07E-06***
23	49	20	0,41	0,7384	0,7414	82	1,67	0,4049	0,3980	83	1,69	7,9E-16***	1,06E-19***
24	61	16	0,26	0,1335	0,1124	111	1,82	0,9382	0,9382	199	3,26	0,0054**	0,0041**
25	43	20	0,47	0,3553	0,3718	67	1,56	0,1793	0,1682	118	2,74	5,2E-05***	1,85E-05***
26	48	22	0,46	0,3678	0,3830	74	1,54	0,1322	0,1218	168	3,50	0,1046	0,0978
27	44	14	0,32	0,5079	0,4959	80	1,82	0,9418	0,9418	152	3,45	0,0881	0,0813
28	41	27	0,66	0,0034**	0,0079**	66	1,61	0,2871	0,2771	165	4,02	0,8385	0,8389
29	46	20	0,43	0,5357	0,5449	69	1,50	0,0922	0,0822	166	3,61	0,2247	0,2178
X	100	70	0,70	1,03E-07***	1,74E-06***	191	1,91	0,5610	0,5636	500	5,00	1E-07***	2,96E-07***
Overall	2448	928	0,38		3,683E-05***	4487	1,83		3,40E-07***	9698	3,96		1,36E-47***

Elements detected by each method. Chr: chromosome; n: number; den: density (elements/Mb).

a Single chromosome against the rest of chromosomes; * p<0,05; ** p<0,01; *** p<0,001. Based on Villesen et al (2004).

↑ More ERVs than expected; ↓ Fewer ERVs than expected.

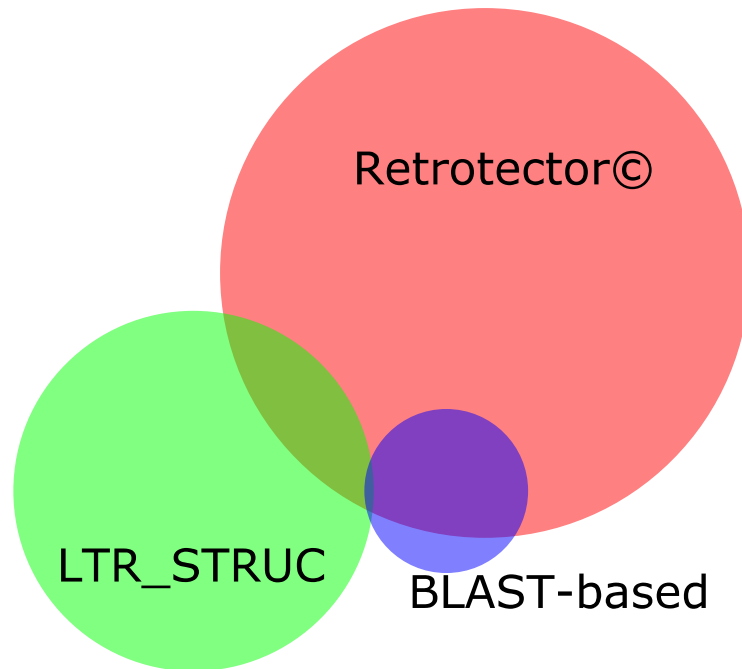


Figure II-1: Diagram representing the number of detected elements (circle size) and common elements (overlapping area).

The gene composition of the elements detected differed from one method to another (Table II-3): The elements detected by Retrovector© had the three major genes but, surprisingly, most of the elements detected by LTR_STRUC did not have the *pol* gene. Most of the BLAST-based search elements were made up of two genes.

The genomic distribution of the ERVs detected through the cow genome was not concomitant with the 3 methods (Table II-2 and supplementary Figures II-1 - II-3). In the case of the BLAST-based search, chromosomes 18, 28 and X had significantly more elements than would be expected in a homogeneous distribution, and chromosomes 14 and 20 had fewer elements. The number of elements detected by LTR_STRUC in chromosomes 1, 2 and 3 was significantly higher than would be expected in a homogeneous distribution, but there were fewer elements than expected in chromosomes 17, 18, 19 and 22. Finally, with Retrovector© chromosomes 1, 2, 3, 6, 9 and X had more elements than expected and chromosomes 13, 18, 19, 22, 23, 24 and 25 had fewer than expected. In these analyses the significance levels were similar with the two statistics used, χ^2 and G-test. Using the G-test the overall distribution of ERVs in the whole genome was found not to be significantly homogeneous in the three methods.

Table II-3: Structure of ERV elements detected by each method.

Structure	Retrovector©	LTR_STRUC	BLAST	Overall ^a
LTR - RT - LTR	5006 (51,61%)	639 (14,24%)	427 (46,01%)	5254 (38,57%)
<i>gag</i> - <i>pol</i> - <i>env</i>	7466 (76,99%)		162 (17,46%)	
<i>gag</i> - <i>pol</i>	2002 (20,64%)		267 (28,77%)	
<i>gag</i> - <i>env</i>	91 (0,94%)		190 (20,47%)	
<i>pol</i> - <i>env</i>	99 (1,02%)		309 (33,30%)	
<i>gag</i>	18 (0,19%)			
<i>pol</i>	22 (0,23%)			
<i>env</i>	0 (-)			
Pol presence	7582 (78,18%)	383 (8,54%)	738 (79,53%)	
Pol absence	2116 (21,82%)	4104 (91,46%)	190 (20,47%)	
Total	9698	4487	928	13622

In parenthesis the percentage of total detected elements.

^a non-redundant ERVs

Moreover, the number of ERVs detected showed a strong correlation with chromosome size in all three methods (BLAST, Spearman's $\rho=0.7677$, $p<0.001$; LTR_STRUC, $\rho=0.9720$, $p<0.001$; and Retrorector©, $\rho=0.9735$, $p<0.001$). In addition, there was a negative correlation between the number of ERVs detected and GC content (BLAST, $\rho=-0.4968$, $p<0.01$; LTR_STRUC, $\rho=-0.7066$, $p<0.001$; and Retrorector©, $\rho=-0.6979$, $p<0.001$). Finally, there was no significant correlation between the number of ERVs detected and the gene density of the chromosomes.

II.4.2. Classification of bovine endogenous retroviruses

A phylogenetic tree based on the well-conserved *pol* gene RT' region of selected BoERVs (detected with at least two elements and having an RT' region with >500 nucleotides) with other

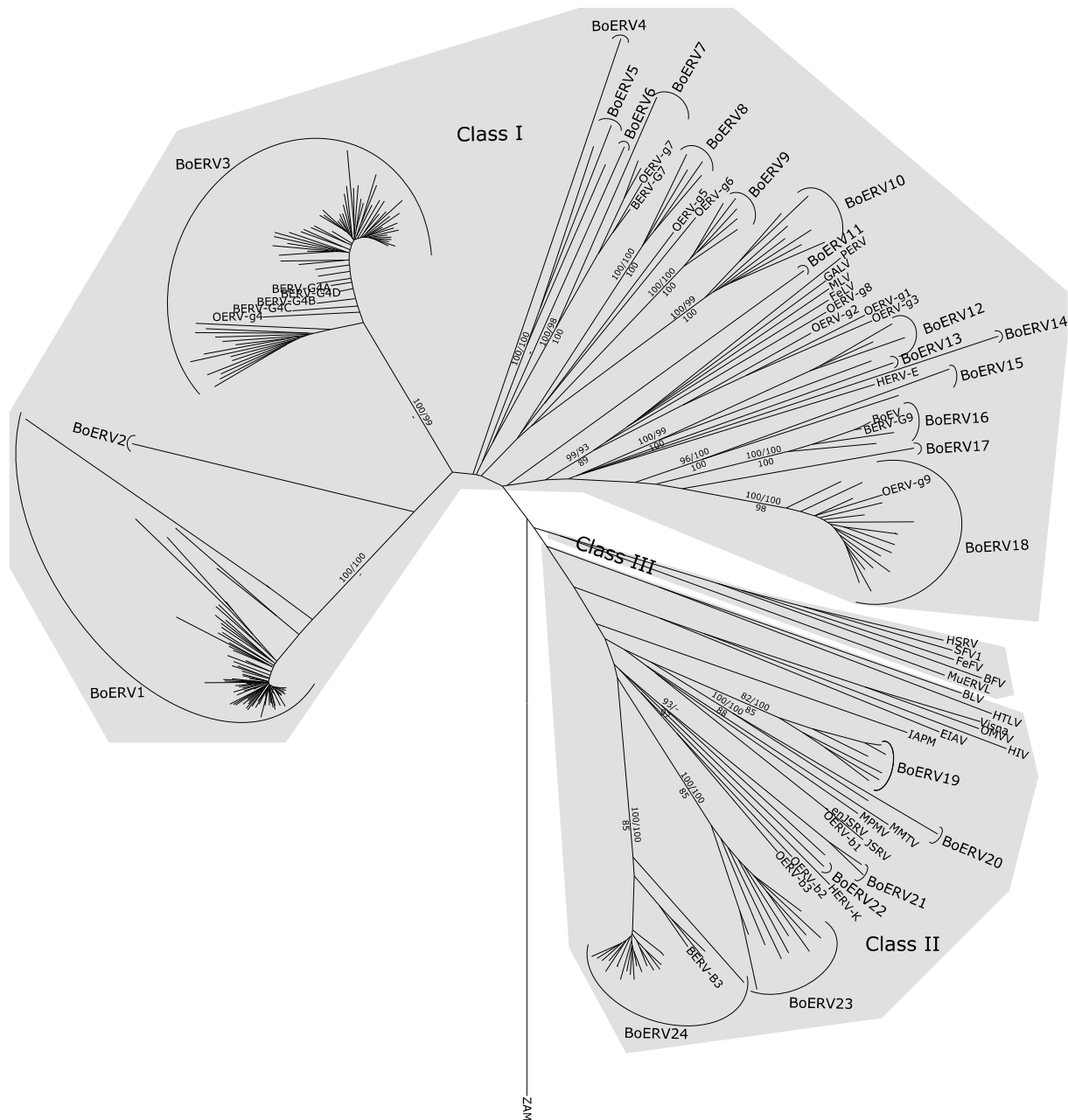


Figure II-2: RT region based phylogenetic tree of BoERVs.

In this tree BoERVs detected by at least 2 methods and with *pol* gene longer than >500 nucleotides were included. Topology was based on Neighbor-Joining method, p-distance, 1000 bootstrap. The tree was rooted with *Drasophila* ZAM element. Above the branches the NJ bootstrap values and ML bootstrap values are showed; below the branches Bayesian Posterior Probability.

endogenous and exogenous retroviruses showed that most of the elements were related to class I or class II outgroup elements, so they can be classified as such by homology. Moreover, there were no class III related elements among the elements analyzed (Figure II-2).

Based on this tree the BoERV elements were classified in families according to the tree topology and the statistical support of the clustering relationships. Overall, 18 families related to class I ERVs and 6 related to class II were defined. The ERV elements detected experimentally in the cow genome were classified within these families, and some of these groups were also related to ovine ERVs. It proved possible to classify all 24 families in three types based on the number of ERVs included (Table II-4). Class I BoERV1, BoERV3 and BoERV18, and Class II BoERV24 were the

Table II-4: Properties of BoERV families characterized in this work.

	Detected (Number)	Length (bases) Media Range	Representative element ^a		Insertion time (MYA)		Nomenclature by Xiao, 2008	Ovine Endogenous Retrovirus Orthologue
Family			PBS	Chr: Start-end (Strand)	Oldest	Youngest		
Class I								
BoERV1	82	10313 7333-16256	Pro	2:69350995-69361048 (-)	209-96	0		
BoERV2	1	10688	-	4:94489122-94499810 (+)				
BoERV3	66	9776 7437-15164	His	7:55058524-55069351 (-)	194-89	7-3	BERV-G4	OERV-G4
BoERV4	1	9917	-	24:39795592-39805509 (+)	51-24			
BoERV5	3	9859 7900-10850	-	18:12944241-12952141 (+)	67-31	65-30		
BoERV6	1	10869	-	3:10016186-10027055 (+)	27-12			
BoERV7	3	10671 10193-11172	Tyr	19:36194623-36204816 (+)	66-30	2-1	BERV-G7	OERV-G7
BoERV8	4	10077 8583-11079	-	5:43721672-43731333 (+)	76-35	37-17		
BoERV9	5	10764 10023-11254	Phe	5:42333382-42344624 (+)	56-26	34-16		
BoERV10	7	10456 7509-12994	Ser	3:97148459-97158193 (-)	205-94	29-13		
BoERV11	1	10947	-	2:123894988-123905935 (+)				
BoERV12	4	9949 7058-11208	Ser	9:6107380-6115865 (-)	82-38	29-13		
BoERV13	1	11087	-	14:26972828-26983915 (-)				
BoERV14	1	9859	-	23:27514449-27524308 (+)	57-26			
BoERV15	2	10052 9908-10196	Pro	23:27654628-27664824 (+)	35-16	35-16		
BoERV16	3	15219 11234-23004	Pro	13:77909792-77919633 (-)	38-17	38-17	BERV-G9	
BoERV17	1	11106	Tyr	18:49897632-49908738 (+)				
BoERV18	16	8209 7428-10311	Pro	9:9401110-9409616 (+)	193-89	10-5		OERV-G9
Class II								
BoERV19	7	10354 8670-13275	His	4:33488089-33501364 (+)	28-13	0		
BoERV20	2	9587 9531-9643	Met	10:23209873-23219516 (-)	27-13	27-12		
BoERV21	3	8881 6283-10345	Lys	2:19509989-19520334 (+)	32-15	15-7		
BoERV22	1	11077	-	18:42572717-42583794 (-)	62-29			
BoERV23	10	10085 7876-11307	-	7:6319253-6327129 (-)	54-25	20-9		
BoERV24	22	10703 8801-25030	Lys	4:69860298-69869823 (-)	212-98	11-5	BERV-B3	
The families were defined with BoERVs detected by, at least, 2 method and their pol gene was >500 nucleotides. In the case of families with one member only one insertion time and the media of the length are showed. In some representative elements the PBS could not be predicted.								

^a The families were defined with BoERVs detected by, at least, 2 method and their pol gene was > 500 nucleotides. In the case of families with one member only one insertion time and the media of the length are showed. In some representative elements the PBS could not be predicted.

most abundant families, with 82, 66, 16 and 22 clustered elements respectively. A second group of families had between 4 and 10 elements. Finally, 8 elements included in the analysis were isolated and had no significant relationship with any other element.

II.4.2.1. Class-I families (BoERV1 to BoERV18)

As mentioned above, most of the BoERVs included in the phylogenetic analysis were related to Class I elements and were classified in 18 families.

The length of Class-I related families were between 8209 bases (BoERV18 family) and 15219 (BoERV16). The longest ERV was from the BoERV16 family (23004 bases) and the shortest from BoERV12 (7058). The length of the longest ERV was due to a duplication of *pol* and *env* retroviral genes. The remaining ERVs showed a typical length of between 7 and 12kb (Boeke & Stoye, 1997).

As far as the most abundant families were concerned, the BoERV1 family was quite heterogeneous. In contrast, the elements of the BoERV3 and BoERV18 families were similarly divergent. Moreover, the BoERV3 family was related to previously described cow BERV-g4 and sheep OERV-g4 elements and BoERV18 was related to the ovine OERV-g9 element.

The representative element of families BoERV1 and BoERV3 had the LPQG and YVDD motifs (Figure 3). In the case of family BoERV18 the first motif was PPQG.

The rest of the groups were represented by very few or by solitary elements. The divergence level varied from one family to another: the elements of BoERV8, BoERV9 and BoERV10 were less divergent than the elements of BoERV5, BoERV7 and BoERV15. BoERV7 was related to the previously detected cow BERV-g7 and ovine OERV-g7 elements, and BoERV16 with the previously detected BoEV (accession number X99924) and BERV-g9 bovine elements. In BoERV12 and BoERV15 the two motifs were conserved. In the representative element of BoERV9 the YVDD motif was present. However, the representative elements of the BoERV2, BoERV4, BoERV5, BoERV6, BoERV10, BoERV11, BoERV16 and BoERV17 groups did not keep, at least one of the two characteristic motifs.

Finally, the representative BoERVs from families with 1 or 2 members had different conservation levels of their functional motifs. BoERV2 had a deletion in the YVDD motif and BoERV4, BoERV13, BoERV14 and BoERV17 had an aminoacid change. BoERV6 and BoERV11 did not conserve either of the motifs.

BoERV1 was the oldest of the Class-I related families because it had the oldest ERV, which was inserted between 209 and 96 million years ago. This family had the most recent insertion activity, since the youngest member was inserted recently. In addition, we were able to amplify a BoERV1 conserved sequence from the RT region which had a length of 150 bases, in sheep samples (data not shown). The youngest family inserted was the BoERV6 family, which was inserted between 27 and 12 million years ago.

II.4.2.2. Class-II families (BoERV19 to BoERV24)

The Class II elements were grouped in 6 families. The range of lengths of Class-II families was between 8881 and 11077 bases. The longest ERV was from BoERV24 (25030 bases) and the shortest from BoERV21 (6283). As with the longest Class-I related ERV, the longest ERV in Class-II related families showed duplicated genes.

The within-group divergence varied: the elements of BoERV23 were more divergent than those of BoERV19 and in the BoERV24 group there were some tightly closed elements and some very divergent ones. The BoERV24 family was related to the previously detected bovine BERV- β 3 element.

The representative elements of all Class-II related families conserved the LPQG and YMDD motifs, except the BoERV23 family whose motifs were different. In this family, the LPQG motif was replaced by QPQG and YMDD by YLDG.

Genome-wide detection and characterization of endogenous retroviruses in Bos taurus

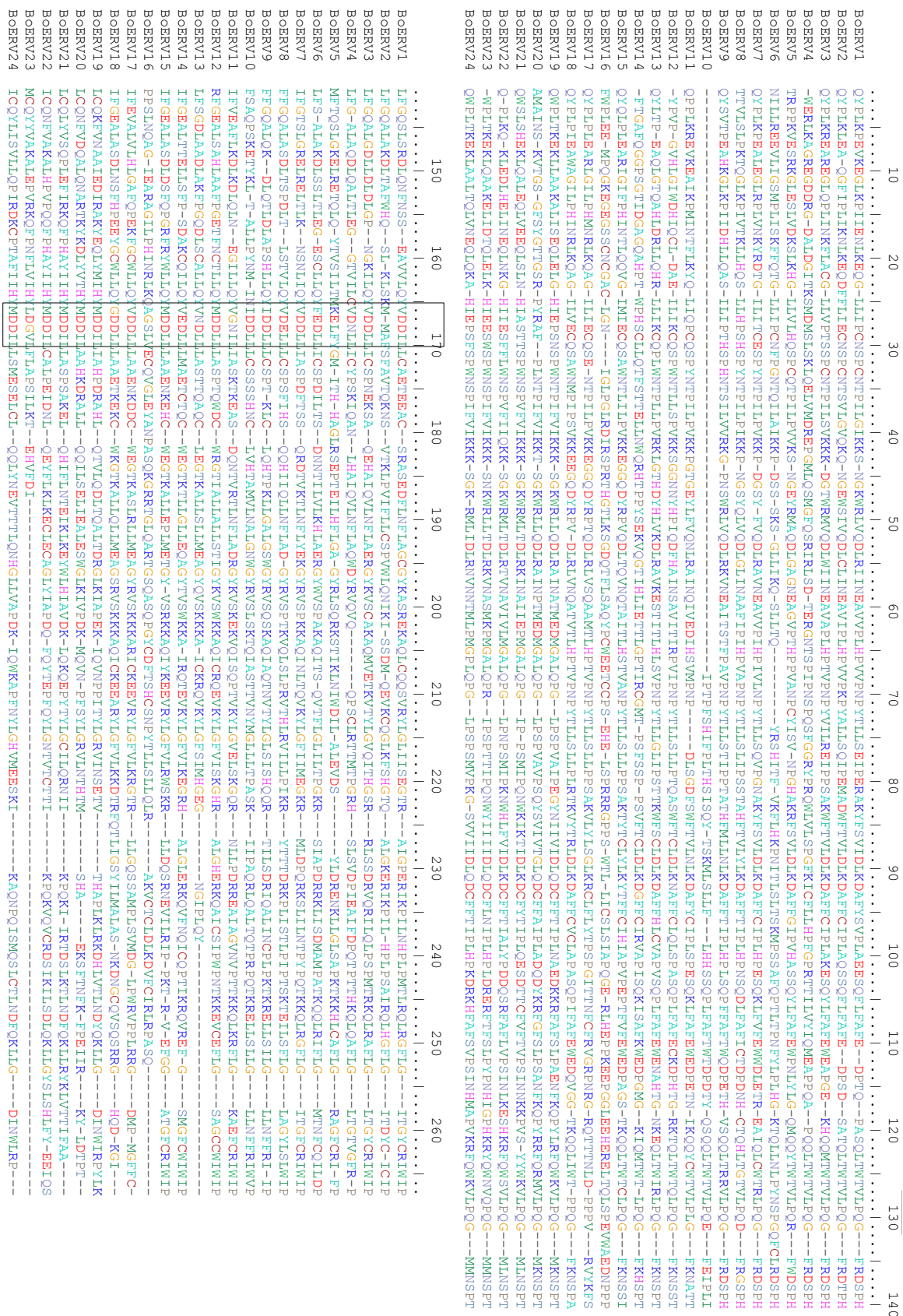


Figure II-3: Partial amino acid sequence of the RT region of the representative ERVs from each putative family. Functional motifs are squared.

The oldest Class-II related family was BoERV24, whose oldest ERV was inserted between 212 and 98 million years ago. The newest families were BoERV19 (between 28 and 13 million years ago) and BoERV20 (between 27 and 13 million years ago). In addition, a member of BoERV19 family was the youngest element, with a recent insertion.

II.4.3. Relationship of BoERVs with other species ERVs

In the phylogenetic tree for the ERVs from the different species, the different ERV elements were grouped in the three classes. In general, within each class, the different representative ERVs clustered following the relationship of their host genome. There was a close relationship between ERVs from cow and sheep, which were clustered together in four phylogenetic lineages (Figure II-4).

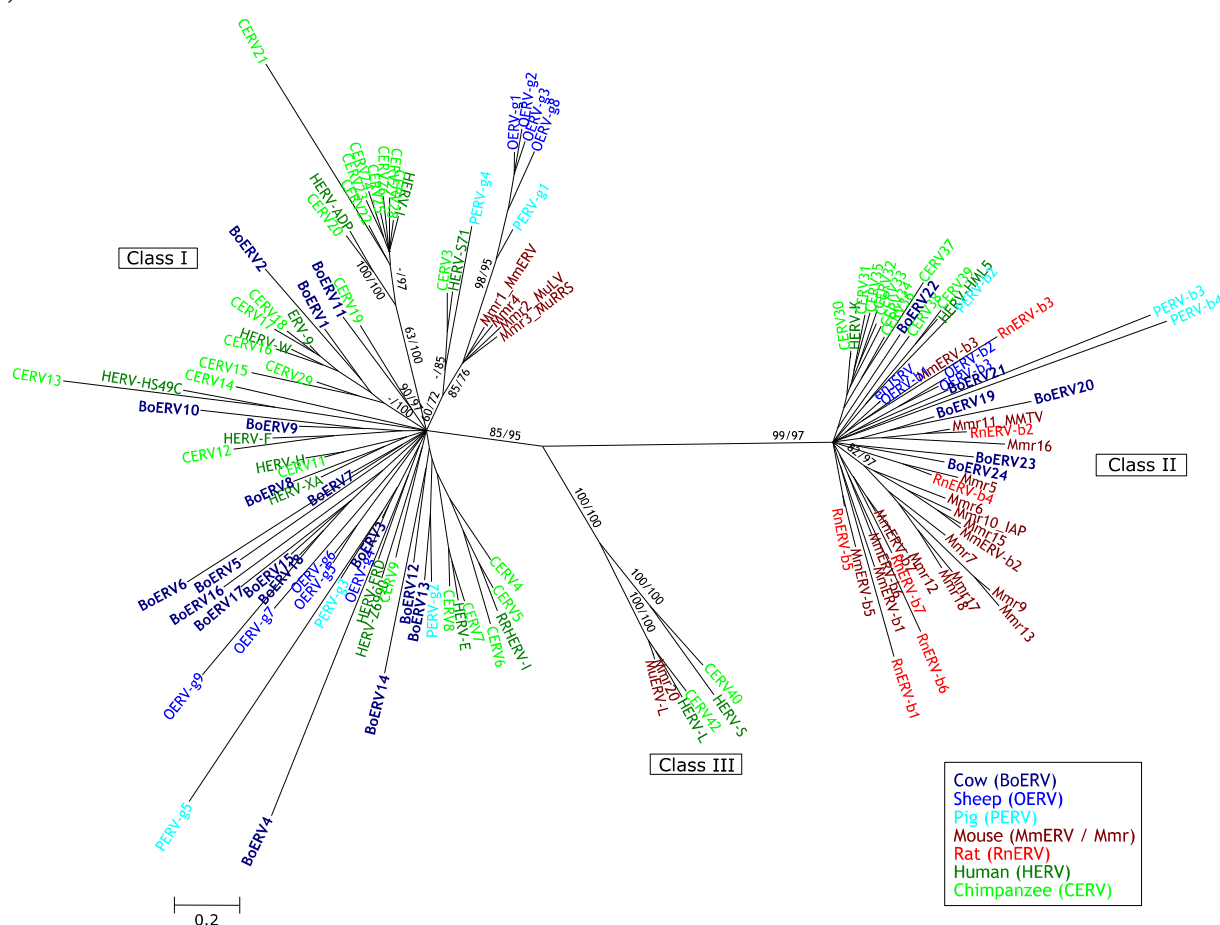


Figure II-4: RT region based unrooted phylogenetic tree of ERVs from different species.

Topology is based on Bayesian inference, 106 generations. In the branches on the left the Maximum Likelihood bootstrap value is shown and on the right the Bayesian Posterior Probability.

The representatives of the different class were grouped in polytomic nodes. However, there were clear relationships between human/chimp families (HERV-I/-ADP and CERV20/21/22/23/24/25), human/chimp/bovine families (ERV9/HERV-W, CERV15/16/17/18/29 and BoERV1/2), chimp/bovine families (CERV19 and BoERV11) and human/chimp/mouse families (HERV-L, CERV42 and MuERV-L/Mmr20). Although the support was not strong, there were bovine/ovine relationships (BoERV15/16/17/18 and OERV-G9) and human/chimp/porcine/bovine (RRHERV-I/HERV-E, CERV4/5/6/7, PERV-g4 and BoERV12/13/14) (see Figure II-4).

Surprisingly, there was one phylogenetic lineage with elements from human, chimpanzee, pig and sheep species, but not cow. This lineage was studied in depth and the insertion-time was estimated using LTR divergence: the human HERV-S71 element was inserted 8.9-19.5 MYA and the

chimpanzee CERV3 element 15.8-33 MYA.

II.5. Discussion

In this work we set out to detect endogenous retroviruses in the cow genome, so the first point to emphasize is the performance of the detection. These results could be an underestimation of the number of possible elements because we used only located genomic information, leaving contigs untested, and because of limited knowledge of these elements, which focused our detection on well-conserved elements. However, this first *in silico* approach is useful for comparing and evaluating the different methods used in recent years and the methods available to detect these elements.

Each method used to detect elements clearly gives us different, and in some cases discordant, information. When the structure of the elements detected was compared and the overlapping detections were calculated some differences appeared.

LTR_STRUC detected more elements without the RT region than with it, in contrast to the other two methods; similarly LTR_STRUC detected many elements which were not detected by other methods. This could be because this detection method is designed to find elements flanked by LTRs, so it was able to detect elements with a non-canonical structure (McCarthy & McDonald, 2003).

BLAST-based search was the method which detected least elements, most of which were also detected by Retrorector©. As the criterion used in BLAST-based search was quite strict, these elements detected could be considered as very well-conserved.

Retrorector© was the method which detected most elements and had most overlapping detections. In most of the elements that it detected the three main genes were identified. It is thus clear that it is more efficient than BLAST-based detection and able to detect less well-conserved elements (Sperber et al., 2007).

The comparison of different genomes was difficult because different methods were used for the detection of ERVs. In human (International Human Genome Sequencing Consortium, 2001), mouse (Mouse Genome Sequencing Consortium, 2002), rat (Rat Genome Sequencing Project Cons, 2004), dog (Lindblad-Toh et al., 2005), cat (Pontius et al., 2007) and cow (Elsik et al., 2009) Repeatmasker and Repbase were used to detect repetitive elements. As stated previously, Repeatmasker and Retrorector© results could not be compared due the length of the elements detected (Sperber et al., 2007). The problem with the length is common to all three methods used in this study: Repeatmasker can detect elements with <1000 bases (Sperber et al., 2007), but the methods used in this study needed longer elements. In the cow genome 142096 ERVs were detected with PALS/PILER (Elsik et al., 2009), 928 by BLAST, 4487 by LTR_STRUC and 9698 by Retrorector©. Surprisingly, the genome coverage of the elements detected given by the different programs was discordant: 1.75% of the genome by PALS/PILER, 0.36% by BLAST, 1.77% by LTR_STRUC and 4.29% by Retrorector©; according to these numbers the coverage was similar or greater with fewer elements. So the slant of short elements in methods such as Repetmasker and PALS/PILER made comparison difficult. In addition, the classification of the elements detected by different programs adds complexity to the comparison: Repeatmasker uses the Repbase annotation (Jurka, 2000; Smit et al., 2004), while in the BLAST-based method a phylogenetic classification is used and Retrorector© uses its own motif database (Sperber et al., 2007). Thus, we found that Repetmasker and the other two methods did not classify the same element in the same class. For example, from ERVs classified as Class I in the BLAST-based method and Retrorector© 64.72% were classified as ERV1 and 35.28% as ERVL by Repeatmasker.

The origin of the different chromosome distributions could lie in the difference in target elements of the different methods. In the analysis of the chromosomal distribution of the elements detected the different methods showed different chromosomes which did not follow any homogeneous distribution. The different nature of the elements detected in each method could be a

good reason for this discrepancy.

Chromosome by chromosome, in the BLAST-based method and in Retrorector© search the X chromosome showed significantly more elements than in a homogeneous distribution. In humans this excess of elements in the X chromosome was also observed (Villesen et al., 2004).

There was a positive correlation between the number of ERVs detected and chromosome length ($p < 0,001$ in all three methods), and a negative correlation with the GC content of the chromosome. However, there was no correlation between the number detected and gene or pseudogene density. In humans it was found that Class I and III ERVs had a negative correlation with GC content, but Class II showed no correlation with GC content (Medstrand et al., 2002). So these insertion preferences of ERVs in the cow genome need to be analyzed in more depth to gain a better understanding of the preferences of bovine ERVs.

Phylogenetic analysis based on the RT region of a number of selected elements showed that these elements could be clustered in 24 putative families, which were called BoERV families. Previously, 4 retroviral families were detected (Xiao et al., 2008b) and included in the 24 putative families detected in this work. Although it had previously been suggested that the BERV-g4 family, referred to here as BoERV3, was the most abundant (Xiao et al., 2008b), we found that BoERV1 was actually the most abundant. This family had not previously been identified in any mammal. We were able to use PCR to amplify a 150 base sequence in sheep, so it could be a ruminant specific ERV family.

The comparison of ERV family number was limited to 4 species with defined families (human, chimpanzee, mouse and rat). In cows, the family number (24 putative families) was greater than for the mouse (20 families) (McCarthy & McDonald, 2004) and smaller than for the chimpanzee (42 families) (Polavarapu et al., 2006) and human (31 families) (Katzourakis et al., 2005). In the case of rodents, where information is only available for Class II elements in two species, the number of families for the cow (6 families) was similar to that for the rat and mouse (7 families) (Baillie et al., 2004). Unfortunately, as far as we know, there is no information on dog and cat ERV families.

No Class III related elements were detected in this study. This lack of Class III related elements could be an artefactual observation due to the distance from the reference sequences used for the search or, more probably, because the presence of Class III ERVs the cow genome is very limited. In fact, although a number of sequences related with Class III were amplified by Bénit et al (Benit et al., 1999), the amplification signal was very weak and these sequences were very short.

The relationship between the representatives of the ERV families of the different species is an interesting point. In general, the lineages of the different ERV groups are divided following the species phylogeny, with humans and chimpanzees together on one side and cows with sheep and pigs on the other side. In rodents only II Class related elements are characterized, and the groups are not so well defined. Even so, representative elements of the human-chimp groups and, to a lesser extent, mouse-rat and pig-sheep groups tend to follow the pattern of previous comparisons of each pair (Klymiuk et al., 2003; Baillie et al., 2004; Polavarapu et al., 2006). Following this pattern the representative bovine elements cluster with the representatives of sheep as obtained by experiment (Klymiuk et al., 2003) in most of the lineages. Interestingly, in one lineage representatives of human, chimpanzee and also pig and sheep groups were present, but there were no cattle elements. In order to solve this absence in cattle, the insertion time of the elements in this lineage was estimated. As there is no genomic information available for pigs and sheep, we could only estimate the insertion times of the human element (8,9-19,5 MYA) and chimpanzee element (15,8-33 MYA). This estimation gave an insertion time later than the divergence of ruminants and primates. Based on the weak support of the tree topology, a single infection is unlikely. In this lineage two independent infections by a similar virus could have been detected and, in the case of ruminants, cows lost this element.

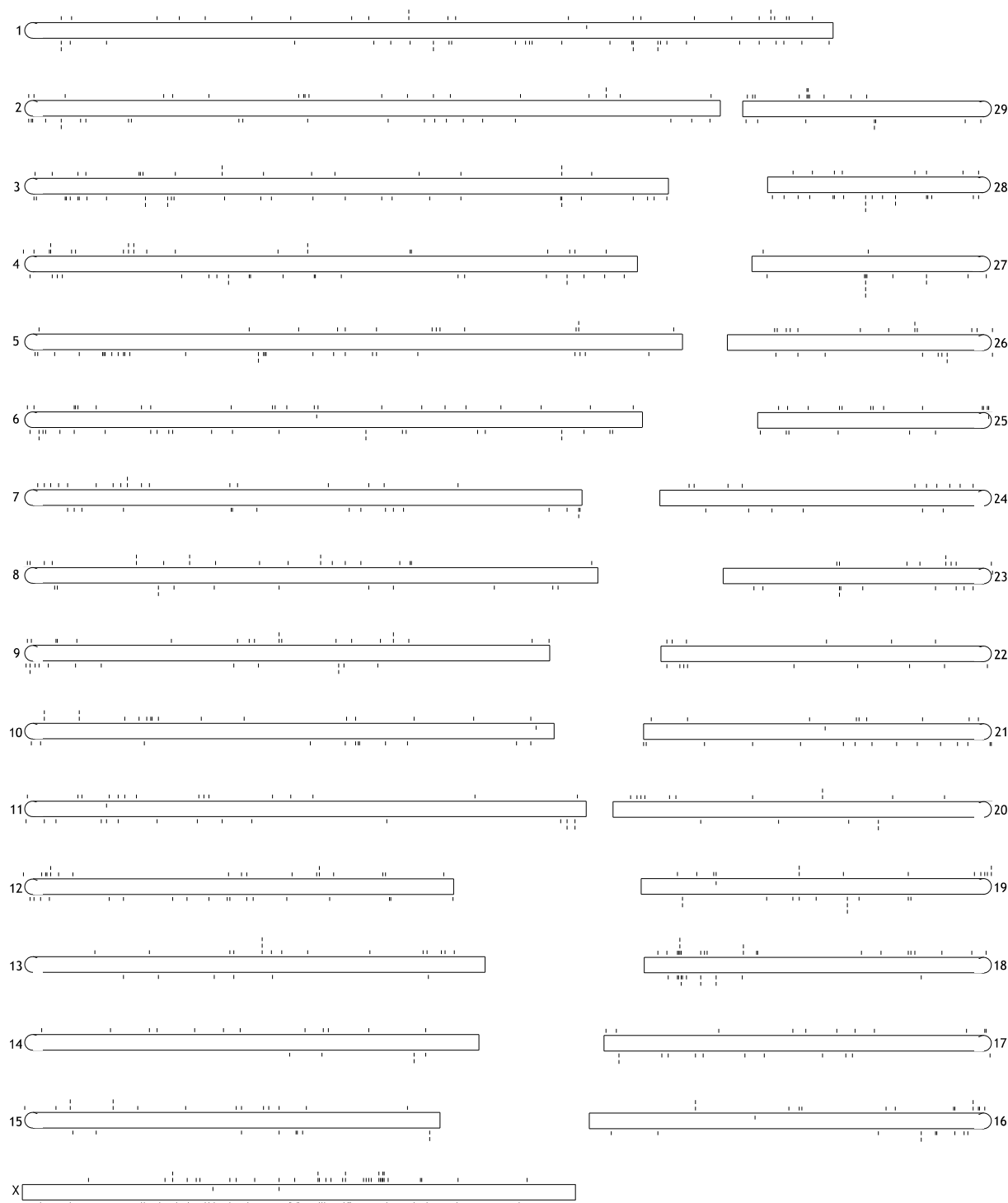
Based on the lack of some families, compared with sheep and pigs, other authors (Xiao et al., 2008b) have suggested that cows have a limited number of ERV families. Taking into account that the number of ERV families described was 31 in humans (Katzourakis et al., 2005), 42 in chimpanzees (Polavarapu et al., 2006), 20 in mice (McCarthy & McDonald, 2004) and 24 putative in cattle (this study), the limitation of ERVs in cow may not be as great as previously stated. Moreover, in this study we detected a family, BoERV1, which had not been detected previously and seemed to be present at least in ruminants. In addition, as pointed out above, we were unable to detect any Class III elements. It has been suggested that in primates and mice (18), ERVs related to this class have been subjected to one or two bursts. Therefore, it is possible that the difference in the number of ERV families with primates and mice could be based on this burst of Class III related ERVs. Finally, the generic picture could be also confused by the selective breeding processes that have accompanied domestication (Boeke & Stoye, 1997). So, although it was true that in cows there could be less ERVs than other species, the difference could be smaller than had been thought.

II.6. References

- Altschul, S., T. Madden, A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D. Lipman. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* **25**:3389-3402.
- Baillie, G., L. van de Lagemaat, C. Baust, and D. Mager. 2004. Multiple groups of endogenous betaretroviruses in mice, rats, and other mammals. *Journal of Virology* **78**:5784-5798.
- Benit, L., J. Lallemand, J. Casella, H. Philippe, and T. Heidmann. 1999. ERV-L elements: a family of endogenous retrovirus-like elements active throughout the evolution of mammals. *Journal of Virology* **73**:3301-3308.
- Blomberg, J., F. Benachenhou, V. Blikstad, G. r. Sperber, and J. Mayer. 2009. Classification and nomenclature of endogenous retroviral sequences (ERVs): Problems and recommendations. *Gene* **448**:115-123.
- Boeke, J. and J. Stoye. 1997. Retrotransposons, endogenous retroviruses, and the evolution of retroelements, p. 343-435. *In* J. Coffin, S. Hughes, and H. Varmos (eds.), *Retroviruses*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Castresana, J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Biol Evol* **17**:540-552.
- Elsik, C., R. Tellam, K. Worley, and Bovine Genome Sequencing Anal Cons. 2009. The Genome Sequence of Taurine Cattle: A Window to Ruminant Biology and Evolution. *Science* **324**:522-528.
- Gifford, R., P. Kabat, J. Martin, C. Lynch, and M. Tristem. 2005. Evolution and distribution of class II-related endogenous retroviruses. *Journal of Virology* **79**:6478-6486.
- Golan, M., A. Hizi, J. Resau, N. Yaal-Hahoshen, H. Reichman, I. Keydar, and I. Tsarfaty. 2008. Human endogenous retrovirus (HERV-K) reverse transcriptase as a breast cancer prognostic marker. *Neoplasia* **10**:521-5U3.
- Guindon, S. and O. Gascuel. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic Biology* **52**:696-704.
- International Human Genome Sequencing Consortium. 2001. Initial sequencing and analysis of the human genome. *Nature* **409**:860-921.
- Jern, P. and J. Coffin. 2008. Effects of Retroviruses on Host Genome Function. *Annual Review of Genetics* **42**:709-732.
- Johnson, W. and J. Coffin. 1999. Constructing primate phylogenies from ancient retrovirus sequences. *Proceedings of the National Academy of Sciences of the United States of America* **96**:10254-10260.
- Jurka, J. 2000. Repbase Update - a database and an electronic journal of repetitive elements. *Trends in Genetics* **16**:418-420.
- Katoh, K., K. Kuma, H. Toh, and T. Miyata. 2005. MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Research* **33**:511-518.
- Katzourakis, A., A. Rambaut, and O. Pybus. 2005. The evolutionary dynamics of endogenous retroviruses. *Trends in Microbiology* **13**:463-468.
- Klymiuk, N., M. Muller, G. Brem, and B. Aigner. 2003. Characterization of endogenous retroviruses in sheep. *Journal of Virology* **77**:11268-11273.
- Kumar, S. and S. Hedges. 1998. A molecular timescale for vertebrate evolution. *Nature* **392**:917-920.
- Kumar, S., K. Tamura, and M. Nei. 2004. MEGA3: Integrated software for molecular evolutionary genetics analysis and sequence alignment. *Briefings in Bioinformatics* **5**:150-163.
- Lindblad-Toh, K., C. Wade, T. Mikkelsen, E. Karlsson, D. Jaffe, M. Kamal, M. Clamp, J. Chang, E. Kulbokas, M. Zody, E. Mauceli, X. Xie, M. Breen, R. Wayne, E. Ostrander, C. Ponting, F. Galibert, D. Smith, P. deJong, E. Kirkness, P. Alvarez, T. Biagi, W. Brockman, J. Butler, C. Chin, A. Cook, J. Cuff, M. Daly, D. DeCaprio, S. Gnerre, M. Grabherr, M. Kellis, M. Kleber, C. Bardeleben, L. Goodstadt, A. Heger, C. Hitte, L.

- Kim, K. Koepfli, H. Parker, J. Pollinger, S. Searle, N. Sutter, R. Thomas, C. Webber, E. Lander, and Broad Inst Genome Sequencing Plat. 2005. Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature* **438**:803-819.
- McCarthy, E. and J. McDonald. 2003. LTR_STRUC: a novel search and identification program for LTR retrotransposons. *Bioinformatics* **19**:362-367.
- McCarthy, E. and J. McDonald. 2004. Long terminal repeat retrotransposons of *Mus musculus*. *Genome Biology* **5**.
- Medstrand, P., L. van de Lagemaat, and D. Mager. 2002. Retroelement distributions in the human genome: Variations associated with age and proximity to genes. *Genome Research* **12**:1483-1495.
- Mouse Genome Sequencing Consortium. 2002. Initial sequencing and comparative analysis of the mouse genome. *Nature* **420**:520-562.
- Patience, C., W. Switzer, Y. Takeuchi, D. Griffiths, M. Goward, W. Heneine, J. Stoye, and R. Weiss. 2001. Multiple groups of novel retroviral genomes in pigs and related species. *Journal of Virology* **75**:2771-2775.
- Perron, H., C. Bernard, J. B. Bertrand, A. B. Lang, I. Popa, K. Sanhadji, and J. Portoukalian. 2009. Endogenous retroviral genes, Herpesviruses and gender in Multiple Sclerosis. *Journal of the Neurological Sciences* **286**: 65-72.
- Polavarapu, N., N. Bowen, and J. McDonald. 2006. Identification, characterization and comparative genomics of chimpanzee endogenous retroviruses. *Genome Biology* **7**.
- Pontius, J., J. Mullikin, D. Smith, K. Lindblad-Toh, S. Gnerre, M. Clamp, J. Chang, R. Stephens, B. Neelam, N. Volfovsky, A. Schaffer, R. Agarwala, K. Narfstrom, W. Murphy, U. Giger, A. Roca, A. Antunes, M. Menotti-Raymond, N. Yuhki, J. Pecon-Slaterry, W. Johnson, G. Bourque, G. Tesler, S. O'Brien, Agencourt Sequencing Team, and NISC Comparat Sequencing Prog. 2007. Initial sequence and comparative analysis of the cat genome. *Genome Research* **17**:1675-1689.
- Posada, D. and K. Crandall. 1998. MODELTEST: testing the model of DNA substitution. *Bioinformatics* **14**:817-818.
- R Development Core Team. 2008. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
- Rat Genome Sequencing Project Cons. 2004. Genome sequence of the Brown Norway rat yields insights into mammalian evolution. *Nature* **428**:493-521.
- Rice, P., I. Longden, and A. Bleasby. 2000. EMBOSS: The European molecular biology open software suite. *Trends in Genetics* **16**:276-277.
- Ronquist, F. and J. Huelsenbeck. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**:1572-1574.
- Serafino, A., E. Balestrieri, P. Pierimarchi, C. Matteucci, G. Moroni, E. Oricchio, G. Rasi, A. Mastino, C. Spadafora, E. Garaci, and P. S. Vallebona. 2009. The activation of human endogenous retrovirus K (HERV-K) is implicated in melanoma cell malignant transformation. *Experimental Cell Research* **315**:849-862.
- Smit, AFA, Hubley, R, and Green, P. RepeatMasker Open-3.0. 2004.
- Ref Type: Internet Communication
- Sokal, R. and F. Rohls. 1969. Biometry: the principles and practise of statistics in biological research. W. H. Freeman and Co, New York.
- Sperber, G., T. Airola, P. Jern, and J. Blomberg. 2007. Automated recognition of retroviral sequences in genomic data - RetroTector (c). *Nucleic Acids Research* **35**:4964-4976.
- Thompson, J., D. Higgins, and T. Gibson. 1994. Clustal-W - Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specifil gap penalties and weight matrix choice. *Nucleic Acids Research* **22**:4673-4680.
- Tristem, M. 2000. Identification and characterization of novel human endogenous retrovirus families by phylogenetic screening of the Human Genome Mapping Project database. *Journal of Virology* **74**:3715-3730.
- Villesen, P., L. Aagaard, C. Wiuf, and F. S. Pedersen. 2004. Identification of endogenous retroviral reading frames in the human genome. *Retrovirology* **1**.
- Xiao, R., J. Kim, H. Choi, K. Park, H. Lee, and C. Park. 2008a. Characterization of the bovine endogenous retrovirus beta 3 genome. *Molecules and Cells* **25**:142-147.
- Xiao, R., K. Park, H. Lee, J. Kim, and C. Park. 2008b. Identification and classification of endogenous retroviruses in cattle. *Journal of Virology* **82**:582-587.
- Xiao, R., K. Park, Y. Oh, J. Kim, and C. Park. 2008c. Structural Characterization of the Genome of BERV gamma 4, the Most Abundant Endogenous Retrovirus Family in Cattle. *Molecules and Cells* **26**:404-408.

II.7. Appendix



Supplementary figure II-1: Chromosomal location of ERVs detected by BLAST.
Above chromosomes antisense strand, under chromosomes sense strand.



Supplementary figure II-2: Chromosomal location of ERVs detected by LTR_STRUC.

Above chromosomes antisense strand, under chromosomes sense strand. In red LTR elements where LTR_STRUC has detected the RT region.

Genome-wide detection and characterization of endogenous retroviruses in *Bos taurus*



II-3 Supplementary figure: Chromosomal location of ERVs detected by Retrorector®.

Above chromosomes antisense strand, under chromosomes sense strand. In colours the class assigned by Retrorector® to each ERV: red, class I; green, class II; blue class III; yellow, class I or II; magenta, class I or III; cyan, class II or III; gray, undetermined.

III. Zaldiaren erretrobirus endogenoen (ERV) detekzioa eta sailkapena

III.1. Sarrera

Erretrobirus endogenoak (ERV) ostalariaren hozi-lerroan txertatu diren erretrobirus exogenoen fase probiralak dira, belaunaldietan zehar genoma ostalariaren txertatuta mantendu direnak (Boeke & Stoye, 1997).

ERVen ohiko egitura hiru gene nagusi aurki ditzakegu: *gag*, birusaren proteina estrukturalak kodetzen dituen genea; *pol*, birusaren erreplikazio eta insertziorako beharrezko entzimak kodetzen dituen genea; eta *env*, birusaren estalkiaren gainazaleko glikoproteinak kodetzen dituen genea. 3 gene hauek osotutako zati zentralaren inguruan LTRak (*long terminal repeats*) aurkitzen dira, kontrol eskualdeak direnak, eta bertan promotoreak, areagotzaileak eta poliadenilazio-seinaleak daude kokatuta (Griffiths, 2001).

Erretrobirus endogenoek erretrobirus exogenoekin duten erlazio filogenetikoaren arabera, hiru klasetan banatzen dira: I klasea, *Gammaretrovirus* eta *Epsilonretrovirus* generoekin erlazio dutenak; II klasea, *Alpharetrovirus*, *Betaretrovirus*, *Deltaretrovirus* eta *Lentivirus* generoekin erlazioa dutenak; eta, azkenik, III klasea, *Spumaretrovirus* generoarekin erlazio urruna duten ERVak (Gifford et al., 2005).

ERVak ornodunetan detektatu dira, batez ere. Arrain, anfibio, narrasti eta hegaztiak ikertuak izan badira ere, batez ere mamalioetan ikertu eta karakterizatu dira. Bestalde, erretrobirus exogenoen generoekin erlazionatutako ERVen presentzia talde hauetan ezberdina da. I klasekoak ornodunetan ostalari hein zabala azaltzen dute; batez ere *Gammaretrovirus* generoarekin erlazionatutako ERVak aurkitu dira, *Epsilonretrovirus* generoarekin erlazionatutako ERV gutxi detektatu dira. II klasekoak, aldiz, batez ere hegaztietan (*Alpharetrovirus*ekin lotutako ERVak) eta mamalioetan (*Betaretrovirus*) detektatu dira, aztertutako espezieetan *Deltaretrovirus*ekin erlazionatutako ERVrik ez da detektatu eta *Lentivirus*ekin erlazionatutakoak orain dela gutxi arte ez dira detektatuak izan. Azkenik, *Spumavirus*ekin ahaidetasun filogenetiko urruna duten ERVak hainbat ornodun-taldetan detektatu dira (Tristem, 2000; Gifford et al., 2005; Katzourakis et al., 2007).

Genoma osoen sekuentziak eskuragarri bilakatu diren heinean, ERVen genoma osoko azterketak lanabes bioinformatiko ezberdinen bidez egin dira. Genomak sekuentziatzen direnean batez ere Repeatmasker programa (Smit et al., 2004) erabiltzen da elementu errepikakorrak deskribatzeko, horien artean ERVak. ERVak espresuki bilatzeko lehenengo azterketak gizakian (*Homo sapiens*) egin ziren BLAST algoritmoan oinarrituz (Tristem, 2000; Villesen et al., 2004), eta metodo berdina erabiliz karraskarietan (*Mus musculus* eta *Rattus norvegicus*) egin ziren bilaketak (Baillie et al., 2004). Txinpantzearen (*Pan troglodytes*) genoma aztertzeke (Polavarapu et al., 2006) LTRdun egiturak espresuki detektatzeko garatutako programa, LTR_STRUC deitutakoa (McCarthy & McDonald, 2003), erabili zen, BLASTen oinarritutako metodologiarekin uztartuz. Estrategia berdina saguetan (McCarthy & McDonald, 2004) eta oiloetan (*Gallus gallus*) (Huda et al., 2008) erabili zen. Azkenik, ERVak bilatzeko garatutako programa bat, Retroector© alegia (Sperber et al., 2007), gizaki, txinpantze eta beste hainbat genomatan ERVak bilatzeko erabili izan da. Gure taldean hiru metodoak, hots, BLAST, LTR_STRUC eta Retroector© erabili dira behiaren (*Bos taurus*) genomaren ERVen bilaketa egiteko [II kapitulua].

Behiaren genomaren analisisan agertu zaizkigun zenbait galderari erantzuteko asmoz, beste abere baten genoma aztertzea planteatu genion geure buruari. Eskuragarri dagoen zaldiaren (*Equus caballus*) genomaren sekuentziazio-mailak, 6.8Xkoa dena (Wade et al., 2009), genoma osoko bilaketetarako aproposa egiten du. Zaldiaren genoma sekuentziatzerakoan, elementu errepikakorrei dagokionez, LTR-gabeko erretrotransposoiaren artean, LINEek genomaren %19.56 osotzen dute eta SINEek %7.01a. ERVak barneratzen diren LTR retrotransposoiek, berriz, %6.27-ko estaldura dute (Wade et al., 2009). Zaldia *Perissodactyla* ordenean kokatzen da eta sekuentziatu den talde taxonomiko honetako lehen espeziea da. Zaldiak, etxekotu den espeziea den heinean, mundu guztian zehar aurki daitezke, baina badira populazio basatiak ere. Euren portaerari dagokionez, taldeak ar menderatzaile

batez, hainbat emez eta euren ondorengoez osotuta daude; arrak heldutakoan bakarti bilakatzen dira euren eme-taldea osatzeko. Zaldia gizakiaren historian zehar ekonomikoki oso garrantzitsua izan den animalia da, garraiorako, lanerako edota aisialdirako erabili baita (Sorin, 2001).

Zaldian hainbat erretrobirus exogeno deskribatu badira ere, hala nola *Spumavirus* generoko *Equine Foamy Virus* (Tobaly-Tapiero et al., 2000), edota *Lentivirus* generoko *Equine infectious anemia virus* (Stephens et al., 1986) delakoak, ERVe buruz gutxi da ezaguna. Erreferentzia bakarretako bat ungulatueta burututako *Southern blot* bidezko lan bat da, zeinean II klaseko ERVak detektatzen saiatu ziren (Hecht et al., 1996), zaldiaren kasuan hibridazio-seinale ahul batzuk lortuz. Bestea, mamaliotan III klasearekin erlazionatutako ERVen gaineko lan batean aurki dezakegu, zeinean *Southern blot* bidez zaldian hibridazioa detektatu zen (Benit et al., 1999). Azken lan honetan zaldian 5 ERVen *pol* geneen sekuentziak lortu ziren.

Hortaz, ERVen ezagumenduan sakontzeko asmotan eta zaldiaren sekuentziazio-maila altua izanda, zaldiaren genomaren egon daitezkeen ERVak lanabes bioinformatikoen bidez detektatzea eta identifikatzea, elementu horiek sailkatzea eta beste espezieetako ERVekin dituzten erlazioak aztertzea dira lan honen helburuak.

III.2. Material eta metodoak

III.2.1. Sekuentzia genomikoa

Lan honetan erretrobirus endogenoak (ERV) bilatzeko zaldiaren genomaren EquCab1 eta EquCab2 bertsioak aztertu ziren. Bi genoma-bertsioak 6.8X mailako sekuentziazioa badute ere, EquCab1 bertsioan sekuentzien %84a kromosometan kokatuak zeuden eta EquCab2 bertsioan sekuentzien %95a. Genomaren bi bertsio hauek Ensembl zerbitzaritik jaitsi ziren (ftp://ftp.ensembl.org/pub/current_fasta/equus_caballus/dna/). EquCab1 genoma-bertsioaren azterketa bukatuta genuenean EquCab2 bertsioa argitaratu zen. Hortaz, EquCab2 bertsioa ere analizatzea erabaki genuen, genoma-bertsio berriek, hots, sekuentziazio hobeei, ERVen bilaketan zer nolako eragina zuten aztertzeko.

III.2.2. Detekzioa

Behian egin zen moduan, zaldian egon zitezkeen ERVak detektatzeko hiru hurbilketa ezberdin erabili eta konparatu ziren.

1) Lehen estrategia sekuentzia-antzekotasunean oinarritutakoa izan zen. Amu-sekuentzia moduan ondo karakterizatutako 12 erretrobirus exogenoen *gag* eta *env* geneen sekuentzia aminoazidikoak zatiak erabili ziren (III-1. taula). *pol* genearen kasuan erretrobirus klase bakoitzaren sekuentzia bat erabili zen; MLV, MPMV (Baillie et al., 2004) eta BFV erretrobirusenak hain zuzen ere. Erabilitako amu-sekuentziak, *pol* genearen kasuan RT (alderantzizko transkriptasa) eskualdearenak izan ziren eta *env* genean TM (transmintza) eskualdearenak. Kromosoma bakoitzean gene hauen zatiak bilatu ziren, NCBI-BLAST 2.2.14 programan inplementatzen den tBLASTn algoritmoa erabilita (Altschul et al., 1997). Horrela, amu-sekuentzia aminoazidikoak sekuentzia itzuliaren aurka bilatu ziren. Bilaketa hauen emaitzak PHPn idatzitako script-en bidez prozesatu eta MySQL datu-base batean gorde ziren.

Balitzko ERVak eraikitzeke gene bakoitzaren emaitzak konparatu ziren. Amu-sekuentziak lortu ziren 12 erretrobirusetan oinarrituta gene arteko distantziak kalkulatu ziren (III-1. taula). Gutxienez distantzia horien barruan bi emaitza positibo kokatzen baziren, eskualdea balitzko ERV baten moduan definitu zen.

2) Bigarren hurbilketan LTRdun elementuak detektatzeko LTR_STRUC (McCarthy & McDonald, 2003) programa erabili zen. LTRdun elementuak bilatzeko programa honek LTR bikoteak bilatzen ditu eta analisi gehigarriak egiten ditu LTRdun elementua konfirmatzeko.

3) Hirugarren estrategian balitzko ERVak detektatzeko Retroector© (Sperber et al., 2007) programa erabili zen. Programa honek eskualde erretrobiralen motibo ezberdinak itx hartuta, gene

erretrobiralak bilatzen ditu eta, horiekin batera, ERVen kokapena determinatzen du.

III-1. taula: Bilaketa genomikoa burutzeko erabilitako erretrobirusak, amu-sekuentzia posizioa eta sekuentzien arteko distantzia.

Sekuentzien arteko distantziak:			Accesion	Amu-sekuentziaren posizioa			Amu-sekuentzien arteko distantziak			
Izena	Generoa	Ostalaria	Number	<i>gag</i>	<i>pol</i>	<i>env</i>	hasiera- <i>gag</i>	<i>gag-pol</i>	<i>pol-env</i>	<i>env-bukatu</i>
I klasea										
GALV	<i>Gammaretrovirus</i>	<i>Hylobates</i>	NC_001885	1255-1888		6998-7555	1255	846	3508	533
MLV	<i>Gammaretrovirus</i>	<i>Muridae</i>	NC_001501	1020-1653	2520-3276	6920-7457	1020	867	3644	875
FeLV	<i>Gammaretrovirus</i>	<i>Felidae</i>	NC_001940	1516-2149		7323-7860	1516	822	3593	588
II klasea										
JSRV	<i>Betaretrovirus</i>	<i>Ovis aries</i>	NC_001494	1347-1667		6502-7111	1347	1444	2830	351
MMTV	<i>Betaretrovirus</i>	<i>Muridae</i>	NC_001503	1181-1774		6786-7362	1181	1194	3077	1443
BLV	<i>Deltaretrovirus</i>	<i>Bos taurus</i>	NC_001414	790-1362		5760-6158	790	572	2874	2261
HTLV	<i>Deltaretrovirus</i>	<i>Homo sapiens</i>	NC_001436	889-1455		5765-6269	889	780	2786	2238
EIAV	<i>Lentivirus</i>	<i>Equus caballus</i>	NC_001450	945-1562		7235-7867	945	807	4107	492
HIV	<i>Lentivirus</i>	<i>Homo sapiens</i>	NC_001802	1035-1401		7727-8261	1035	765	4805	920
Visna	<i>Lentivirus</i>	<i>Ovis aries</i>	NC_001452	973-1566		8176-8806	973	620	5237	396
III klasea										
HSRV	<i>Spumavirus</i>	<i>Homo sapiens</i>	NC_001795	2111-2750		8788-9399	2111	868	4456	2555
BFV	<i>Spumavirus</i>	<i>Bos taurus</i>	NC_001831	2016-2652	3487-4204	8903-9515	2016	835	4699	2487
Bataz beste ± desbideratze estandar							1373±564	868±217	3851±799	1410±962

III.2.3. Banaketa

Elementuen banaketaren homogeneotasuna Villesen et al.-ek (Villesen et al., 2004) proposatu zuten bezala aztertu zen. Itxarotako elementu kopurua (bataz besteko dentsitatean eta kromosomaren luzeran oinarritua) behatutako kopuruarekin konparatu zen; gure kasuan, kromosoma bakoitzeko ERV kopurua gainontzeko kromosomekin erkatu zen. Sendotasun estatistikoa χ^2 eta G-testa (Sokal & Rohls, 1969) erabiliz kalkulatu ziren. Test bietan askatasun-graduen kopurua batekoa izan zen.

Metodo bakoitzak detektatutako elementu kopuruaren, eta kromosoma-tamainaren, GC edukiaren, gene dentsitatearen eta pseudogene dentsitatearen arteko korrelazio-analisiak R lengoaiarekin (R Development Core Team, 2008) burutu ziren.

III.2.4. Sailkapena

Detektatutako ERVen erlazioa *pol* geneko RT eskualdearen analisi filogenetikoaren bidez aztertu zen. Behintzat bi metodoz detektaturiko eta BLAST bilaketetan antzekotasun maila altua (500 nukleotidotik gorakoak) azaldu zuten 64 sekuentzia eta esperimentalki detektatutako zaldiaren 5 ERVak (Benit et al., 1999) erabili ziren analisi hauetan. Taldez-kanpoko moduan aurreko lanetan erabilitako beste erretrobirus batzuk erabili ziren, hala nola, BoEV (X99924), HERV-E (M10976), PERV (AJ293656), MPMV (NC_001550), IAPM (M17551), OMVV (NC_001511), FeFV (U78765), MuERV-L (Y12713), eta ZAM (AJ000387). Sekuentzia nukleotidikoak MAFFT 5.861 (Katoh et al., 2005) programa erabiliz (FFT-NS-1 aukera) lerrokatu ziren eta lerrokatzea Gblocks 0.91 (Castresana, 2000) programa erabiliz garbitu zen (programa honetan erabilitako aukerak: blokearen luzera minimoa: 5; baimendutako hutsuneak: erdiarekin; alboko posizioentzat sekuentzien kopuru minimoa: 58; ez-kontserbatutako aldameneko posizioen kopuru maximoa: 10). Hiru zuhaitz eraiki ziren: MEGA 3.1 (Kumar et al., 2004) programan inplementatzen den Neighbor-Joining metodoa (aukerak: p-distantzia, bikote-ezabapena eta 1000 bootstrap), Phyml 2.4.4 (Guindon & Gascuel, 2003) programan inplementatzen den sinesgarritasun maximoko metodoa (aukerak: 1000 bootstrap, GTR+G eredua, $\alpha=1.98$, Modeltest 3.7 bidez estimatua (Posada & Crandall, 1998)) eta MrBayes 3.1 (Ronquist & Huelsenbeck, 2003) programan inplementatzen den inferentzia Bayestar metodoa (aukerak: 10^6 belaunaldi, eredua finkatu gabe) erabilita.

Zaldiaren ERVen familiak zuhaitz filogenetikoaren topologiaren sendotasunean oinarrituta definitu

ziren. Duela gutxi ERVen izenak estandarizatzeko proposamena egin bada ere (Blomberg et al., 2009) familiak aurreko nomenklatura erabiliz izendatzea lehenetsi genuen. Hau da, beste espezieetan ingelesezko *human* (gizaki), *ovine* (ardi) eta *porcine* (txerri) erabili da ERVak izendatzeko. Beraz, *equine endogenous retrovirus* bezala izendatu ditugu lan honetan detektatutako ERVak eta EqERV akronimoa erabiltzea erabaki genuen.

Familiatan sailkatu ziren ERVen insertzio-denbora estimatzeko, LTR_STRUC edota Retrorector© analisisietatik lortutako LTRak erabili ziren. Estimazioa egiteko gizaki eta txinpantzeetan egindako analisisetan oinarritutako LTRen dibergentzia erabili zen eta lan horretatik ondorioztatutako $2.3 \cdot 10^{-9}$ ordezkapen/leku/urte-ko tasa aplikatu zitzaion insertzio-denbora kalkulatzeko (Johnson & Coffin, 1999).

Familia bakoitzeko ordezkari gisa familiaren akordio-sekuentziatik gertuen zegoen ERVa aukeratu zen edo, akordioa-sekuentziarik egon ezean, stop kodoi gutxien zituen ERVa hartu zen ordezkartizat. Familien akordioa-sekuentzia eraikitzeke familia bakoitzeko kideen sekuentzia aminoazidikoak ClustalW (Thompson et al., 1994) bidez lerrokatu ziren eta akordioa EMBOSS suitekoa den *cons* programaren bidez lortu zen (Rice et al., 2000). Hurbilen zegoen ERVa aukeratu ahal izateko, kide bakoitzaren sekuentziaren eta akordio-sekuentziaren arteko distantziak MEGA 3.1 erabiliz kalkulatu ziren (aukerak: No of differences eta pairwise deletion).

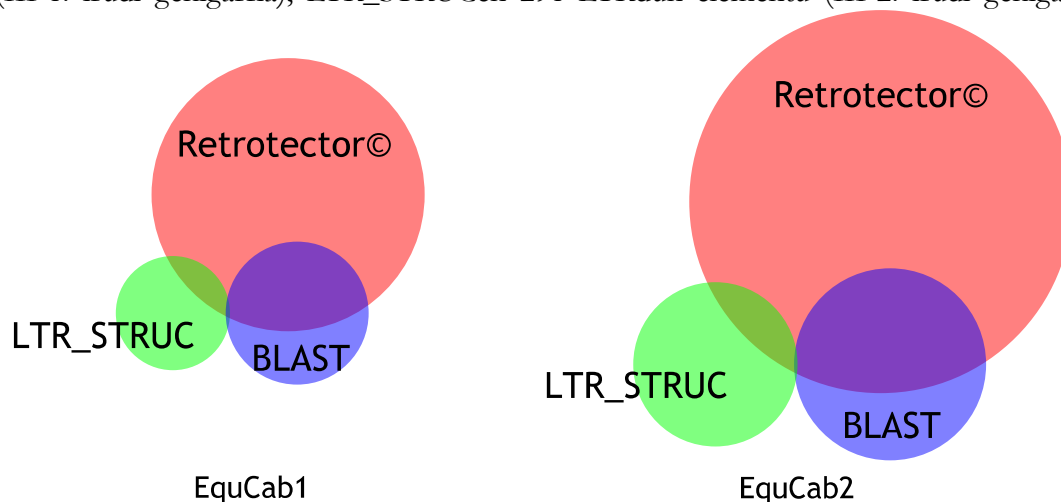
III.2.5. Interespezie-konparaketa

Zaldiaren ERVen familien eta beste espezieen ERVen familien arteko erlazioa aztertzeke zuhaitz filogenetiko bat eraiki zen. Gizakien ERVen 16 familien ordezkari (Tristem, 2000), txinpantzearen 38 (Polavarapu et al., 2006), saguaren 7 (Baillie et al., 2004) eta 17 (McCarthy & McDonald, 2004), arratoiaren 7 (Baillie et al., 2004), ardiaren 14 (Klymiuk et al., 2003), txerriaren 8 (Patience et al., 2001), behiaren 24 [II kapitulua] eta zaldiaren 15 (lan hau) erabili ziren. Sekuentziak MAFFT erabiliz (linsi aukera) lerrokatu ziren eta sekuentzien erdian baino gehiagotan hutsunea agertzen zuten posizioak baztertu ziren. Zuhaitza MrBayes 3.1an inplementatzen den inferentzia Bayestarra erabiliz (10^6 belaunaldi, RtREV matrizea+G+I) eraiki zen.

III.3. Emaizak

III.3.1. Detekzioa

Equus caballus-en genomaren EquCab2 bertsioan detektatutako balizko ERVen kopurua metodoz metodo desberdina izan zen (III-1. irudia): BLASTen oinarritutako bilaketan 378 ERV detektatu ziren (III-1. irudi gehigarria), LTR_STRUCen 291 LTRdun elementu (III-2. irudi gehigarria) eta



III-1. irudia: Aztertutako zaldiaren bi genoma-bertsioetan detektatutako elementu-kopurua (zirkuluen tamainarekiko) eta elementu komunak (gainjartzen den azalera) irudikatzen dituen Venn diagrama.

51 Aztertutako bi bertsioen zirkuluen tamaina proportzionala da, hortaz, bertsio biak erka daitezke.

Retrovector©-ek 1615 balizko ERV (III-3. irudi gehigarria). LTR_STRUC-en kasuan 5. eta 20. kromosoman programak arazoak eman zituen eta, hortaz, 5. kromosoman 2 megabase (47.442.909 eta 49.840.140 posizioen artean) eta 20. kromosoman 5 megabase (27.099.849 eta 32.083.080 posizioen artean) aztertu gabe gelditu ziren.

ERV gutxi batzuk hiru metodoen bitartez detektatuak izan ziren EquCab2 genomaren bertsioan (31 ERV). BLASTen oinarritutako bilaketan detektatutako ERV asko Retrovector©-en bidez ere detektatu ziren (257 ERV). LTR_STRUCek detektatutako LTRdun elementu gutxi batzuk beste metodoek detektatu zituzten (BLAST-en oinarritutakoarekin batera 1, Retrovector©-ekin batera 51). Azkenik, Retrovector©-ek bakarrik detektatutako ERVak asko izan ziren (1276 elementu). EquCab1ekin konparatuz, oro har, detektatutako ERV kopurua txikiagoa izan bazen ere, antzeko proportzioak lortu ziren. (III-1. irudia).

III-2. taula: Zaldia genoma-bertsio bakoitzean detektatutako elementu-kopuruen arteko ezberdintasuna.

Ezb: EquCab2 - EquCab1 arteko ezberdintasuna; %: $(\text{EquCab2} - \text{EquCab1}) * 100 / \text{EquCab1}$.

Kr	Analizatutako luzera		BLAST		LTR_STRUC		Retrovector©	
	Ezb (Mb)	%	Ezb	%	Ezb	%	Ezb	%
1	+6	(+3,33)	+3	(+9,37)	-1	(-5,88)	+12	(+11,11)
2	+15	(+14,15)	+1	(+8,33)	+7	(+100)	+18	(+31,57)
3	+9	(+8,18)	+3	(+27,27)	0		+21	(+38,18)
4	+11	(+11,22)	0		0		+10	(+15,15)
5	+27	(+36,99)	+10	(+142,85)	+8	(+72,72)	+37	(+102,77)
6	+3	(+3,66)	+1	(+8,33)	+5	(+100)	0	
7	+13	(+15,12)	+9	(+60)	+2	(+20)	+31	(+48,43)
8	+7	(+8,05)	-1	(-5,55)	0		+1	(+1,69)
9	+8	(+10,53)	+4	(+25)	-1	(-11,11)	+9	(+29,03)
10	+5	(+6,33)	-1	(-3,70)	-2	(-18,18)	+8	(+11,11)
11	+3	(+5,17)	+3	(+30)	+3	(+30)	+4	(+11,76)
12	+14	(+73,68)	+2	(+40)	+2	(+100)	+18	(+94,73)
13	+26	(+152,94)	+3	(+42,85)	+5	(+500)	+19	(+158,33)
14	+2	(+2,17)	0		+2	(+15,8)	+7	(+13,46)
15	+4	(+4,55)	+1	(+8,33)	+1	(+11,11)	+12	(+25,53)
16	0		0		+2	(+28,57)	+5	(+14,70)
17	+2	(+2,53)	-1	(-10)	-1	(-25)	+2	(+487)
18	+19	(+29,69)	+4	(N/A)	+4	(+200)	+5	(+14,70)
19	+4	(+7,14)	-1	(-12,5)	+2	(+50)	+2	(+7,14)
20	+6	(+10,34)	+2	(+22,22)	+3	(+33,33)	+18	(+36)
21	+5	(+9,43)	+2	(+33,33)	+2	(+66,66)	+7	(+25)
22	0		-1	(-11,11)	+1	(+50)	-5	(-16,12)
23	+15	(+36,59)	0		-1	(-25)	+13	(+86,66)
24	0		-1	(-14,28)	+1	(+33,33)	-3	(-14,28)
25	+5	(+14,29)	+2	(+28,57)	-3	(-50)	+1	(+3,33)
26	+1	(+2,44)	-1	(-20)	+2	(+66,66)	0	
27	+27	(+69,23)	+4	(+66,66)	+6	(+120)	+16	(+76,19)
28	-10	(-17,86)	+1	(N/A)	+1	(+20)	+1	(+7,14)
29	+2	(+6,25)	+1	(+33,33)	+1	(N/A)	+5	(+26,31)
30	+10	(+50)	-1	(-50)	+1	(+100)	+5	(+71,42)
31	0		0		0		-2	(-25)
X	+93	(+300)	+24	(+300)	+38	(+1900)	+148	(+477,41)
Guztira	+332	(+16,08)	+72	(+23,52)	+90	(+44,77)	+425	(+35,71)

EquCab1 eta EquCab2 bertsioak erkatuz, EquCab2 genoma-bertsioan BLASTekin 72 (%23,52 gehiago), LTR_STRUCekin 90 (%44,77) eta Retrovector©-ekin 425 (%35,71) elementu gehiago detektatu ziren 332 Mb gehiagotan (EquCab1 bertsioaren konparatuz, %16,08 gehiago). Hala ere, kromosomaz-kromosoma ezberdintasunak egon ziren, kromosomaren arabera elementu gehiago edo gutxiago detektatu baitziren (III-2. taula)

III-3. taula: Metodo bakoitzak detektatutako elementuak EquCab2 genoma-bertsioan.

Kr.	Aztertutako luzera (Mb)	BLAST				LTR_STRUC				Retrorector®			
		n	den	χ^2 test p-balioa ^a	G-test p-balioa	n	den	χ^2 test p-balioa ^a	G-test p-balioa	n	den	χ^2 test p-balioa ^a	G-test p-balioa
1	186	35	0,19	0,2758	0,2891	16	0,09	0,1458	0,1262	120	0,65	0,6208	0,6185
2	121	13	0,11	0,1531	0,1304	14	0,12	0,8447	0,8435	75	0,62	0,4583	0,4525
3	119	14	0,12	0,2591	0,2382	12	0,10	0,5023	0,4899	76	0,64	0,6323	0,6294
4	109	11	0,10	0,1265	0,1027	10	0,09	0,3580	0,3373	76	0,70	0,7598	0,7610
5	100 ^b	17	0,17	0,7516	0,7545	19	0,19	0,0367*	0,0537	73	0,73	0,4840	0,4896
6	85	13	0,15	0,9105	0,9100	10	0,12	0,9119	0,9115	49	0,58	0,2659	0,2542
7	99	24	0,24	0,0301*	0,0439*	12	0,12	0,9873	0,9873	95	0,96	0,0004***	0,0009***
8	94	17	0,18	0,5641	0,5728	12	0,13	0,8671	0,8681	60	0,64	0,6691	0,6665
9	84	20	0,24	0,0589	0,0784	8	0,10	0,4782	0,4614	40	0,48	0,0247*	0,0179*
10	84	26	0,31	0,0004***	0,0016**	9	0,11	0,6960	0,6901	80	0,95	0,0015**	0,0028**
11	61	13	0,21	0,2696	0,2942	13	0,21	0,0383*	0,0607	38	0,62	0,6243	0,6200
12	33	7	0,21	0,4279	0,4513	4	0,12	0,9928	0,9928	37	1,12	0,0016*	0,0040*
13	43	10	0,23	0,2122	0,2437	6	0,14	0,7360	0,7418	31	0,72	0,7037	0,7069
14	94	10	0,11	0,2012	0,1752	15	0,16	0,2837	0,3055	59	0,63	0,5785	0,5743
15	92	13	0,14	0,6864	0,6812	10	0,11	0,7142	0,7093	59	0,64	0,6990	0,6967
16	87	11	0,13	0,4545	0,4390	9	0,10	0,6181	0,6091	39	0,45	0,0091*	0,0055*
17	81	9	0,11	0,2828	0,2573	3	0,04	0,0262*	0,0094**	43	0,53	0,1110	0,0982
18	83	4	0,05	0,0106*	0,0028**	6	0,07	0,1886	0,1557	39	0,47	0,0213	0,0150
19	60	7	0,12	0,4176	0,3958	6	0,10	0,6247	0,6139	30	0,50	0,0968	0,0820
20	64 ^b	11	0,17	0,7722	0,7753	12	0,19	0,0688	0,0964	68	1,06	0,0001***	0,0004***
21	58	8	0,14	0,7012	0,6951	5	0,09	0,4322	0,4075	35	0,60	0,5090	0,5015
22	50	8	0,16	0,9669	0,9670	3	0,06	0,2059	0,1612	26	0,52	0,1807	0,1633
23	56	5	0,09	0,1921	0,1559	3	0,05	0,1389	0,0965	28	0,50	0,1089	0,0932
24	47	6	0,13	0,6005	0,5882	4	0,09	0,4671	0,4422	18	0,38	0,0142*	0,0076**
25	40	9	0,23	0,2797	0,3095	3	0,08	0,3927	0,3578	31	0,78	0,4315	0,4422
26	42	4	0,10	0,3038	0,2673	5	0,12	0,9995	0,9594	17	0,40	0,0321*	0,0207*
27	66	10	0,15	0,8980	0,8973	11	0,17	0,2890	0,3148	37	0,56	0,2561	0,2424
28	46	1	0,02	0,0190*	0,0033**	6	0,13	0,8648	0,8663	15	0,33	0,0037**	0,0013**
29	34	4	0,12	0,5537	0,5352	1	0,03	0,1201	0,0624	24	0,71	0,8182	0,8196
30	30	1	0,03	0,0843	0,0361*	2	0,07	0,3842	0,3413	12	0,40	0,0660	0,0468*
31	25	5	0,20	0,5924	0,6072	2	0,08	0,5475	0,5211	6	0,24	0,0079**	0,0022**
X	124	32	0,26	0,0039**	0,0078**	40	0,32	0,0000**	0,0000***	179	1,44	0,0000***	0,0000***
Guz	2397	378	0,16	0,0001***	0,0001***	291	0,12	0,0000***	0,0001***	1615	0,67	0,0000***	0,0000***

Kf, kromosoma; n: kopurua; den: dentsitatea (elementu/MB). ^a Kromosoma bat gainontzeko kromosomen aurkai; * p<0,05; ** p<0,01; *** p<0,001
^b LTR_STRUC-en kasuan, 5. kromosoman 98Mb eta 20. kromosoman 59Mb

Metodo bakoitzak detektatutako ERVen gene egitura ezberdina izan zen (III-4. taula). BLASTen oinarritutako bilaketan ERV gehienak bi genez osatuak zeuden; Retrorector®-ek, aldiz, hiru gene nagusiak detektatu zituen. Bi metodo hauetan detektatutako ERV gehienek *pol* genea azaldu zuten baina LTR_STRUCek detektatutako LTRdun elementu gehienek ez zuten gene hau. Bestalde, hiru metodoek detektatutako ERVen ia erdiak LTR-RT-LTR egitura izan zuten, “elementu oso” moduan izendatuak izan direnak (Polavarapu et al., 2006). Azkenik, EquCab2 genoma-bertsioan burututako bilaketan metodo bakoitzak lortutako egituren proportzioak EquCab1-ean lortutakoen antzekoa izan zen (ikus III-1. taula gehigarria).

III-4. taula: Metodo bakoitzak detektatutako elementuen egitura EquCab2 genomaren bertsioan.

Egitura	Retrotector©	LTR_STRUC	BLAST	Guztira ^a
LTR - RT - LTR	865 (%53,56)	39 (%13,40)	-	877 (%45,04)
GAG - POL - ENV	1235 (%76,47)		54 (%14,29)	
GAG - POL	298 (%18,45)		76 (%20,11)	
GAG - ENV	26 (%1,61)		29 (%7,67)	
POL - ENV	38 (%2,35)		219 (%57,94)	
GAG	11 (%0,68)			
POL	4 (%0,25)			
ENV	2 (%0,12)			
Besteak	1 (%0,06)			
<i>pol</i> presentzia	1575 (%97,52)	39 (%13,40)	349 (%92,33)	
<i>pol</i> absentsia	40 (%2,48)	252 (%86,90)	29 (%7,67)	
Guztira	1615	291	378	1947

^a ERVs ez erredundanteak

Detektatutako ERVen banaketa zaldiaren EquCab2 genoma-bertsioan (hemendik aurrerako analisiak egiteko erabiliko dena) ezberdina izan zen metodo bakoitzean (III-3. taula). BLASTen oinarritutako bilaketan 7., 10. eta X kromosomek banaketa homogeen batean itxaron daitekeena baino ERV gehiago zituzten, 18., 28. eta 30. kromosomek, aldiz, gutxiago. LTR_STRUCen kasuan, 5. eta 11. eta X kromosomek itxaron zitezkeenak baino LTRdun elementu gehiago azaldu zituzten χ^2 froga estatistikoan; 17. kromosoman, aldiz, esangarriki elementu gutxiago egon ziren. Azkenik, Retrotector©-en emaitzetan 7., 10., 12., 20. eta X. kromosometan itxarotakoak baino gehiago zeuden eta 9., 16., 24., 26., 28., 30. eta 31. kromosometan gutxiago.

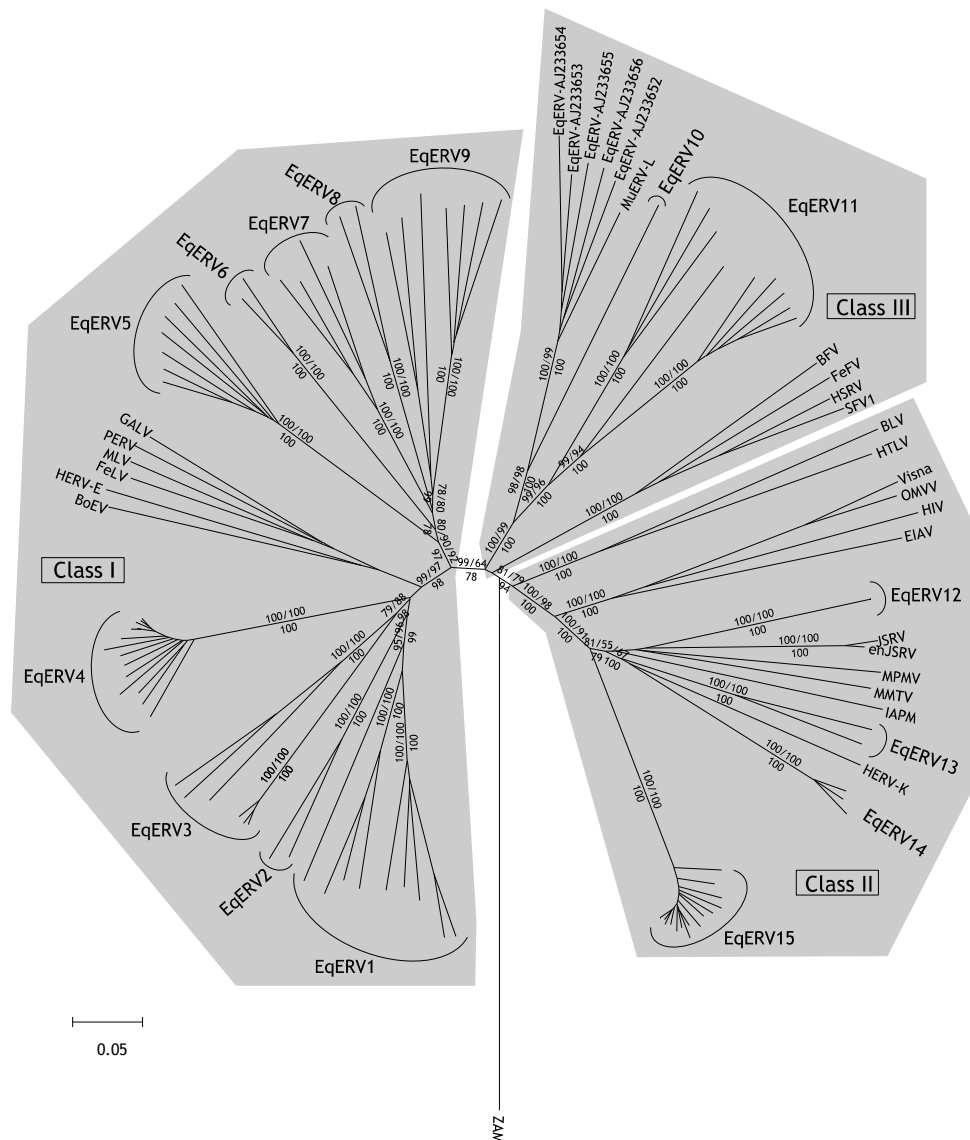
Lehen aipatutako LTR_STRUCen χ^2 bidez detektatutako 5. eta 11. kromosomak izan ezik, banaketa homogeenoa ez zuten gainontzeko kromosometan aldentze hau esangarria izan zen erabilitako bi test estatistikoetan, χ^2 eta G-test frogetan hain zuzen ere. Hiru metodoetan aurkitutako elementuen banaketa, G-test bidez testatua, ez zela esangarriki homogeenoa ondorioztatu zen.

Banaketa ez homogeenoz gain, detektatutako ERVen kopuruek kromosomaren tamainarekin korrelazio sendoa azaldu zuten hiru metodoetan (BLASTen Spearman-en $\rho=0.8093$, $p<0.001$; LTR_STRUCen $\rho=0.8503$, $p<0.001$; eta Retrotector©-en $\rho=0.9068$, $p<0.001$). Bestalde, ERVen kopuruek ez zuten kromosomaren GC edukiarekin, kromosomaren gene dentsitatearekin eta pseudogene dentsitatearekin korrelazio esangarririk azaldu.

III.3.2. Zaldien ERVen sailkapena

Analisi honetan sartu ziren ERVak honakoak izan ziren: behintzat bi metodoz detektatuak izan zirenak eta euren *pol* genea >500 nukleotidoko luzera izan zutenak. Hainbat eretrovirus endogeno eta exogeno eta ondo kontserbatutako elementuen *pol* geneko RT eskualdearekin egindako zuhaitz filogenetikoan oinarrituta detektatutako elementu gehienak I klasekoak zirela ondoriozta daiteke, II eta III klasekoak urriagoak izanda (III-2. Irudia).

EqERV (*Equine endogenous retrovirus*) bezala izendatutako zaldien ERVak, analisi filogenetikoan lortutako zuhaitzaren arabera familietan sailkatu ziren. Metodo filogenetikoaren sendotasunean oinarrituta I klaseko 9 familia, II klaseko 4 familia eta III klaseko 2 familia definitu ziren. Familia hauetatik bat ERV bakartiz osotua zegoen (EqERV10), beste ERVekin erlazio esangarririk gabeko elementua. Talderik handienak EqERV4 eta EqERV15 izan ziren, 14 eta 16 ERVekin bakoitza hurrenez hurren (III-5. Taula).



III-2. irudia: RT eskualdean oinarritutako EqERV-en zuhaitz filogenetikoa.

Topologia Neighbor-Joining metodoan oinarrituta dago (p-distance, 1000 bootstrap). *Drosophilaren* ZAM elementua erro gisa erabili zen. Nodoen gainean NJ eta ML bootstrap balioak erakusten dira eta azpian ondorengo probabilitate Bayestarra.

III.3.2.1. I klasearekin erlazionaturiko familiak. (EqERV1-EqERV9)

Familia hauen gehiengoa hainbat ERVz osotuta zegoen. EqERV1 eta EqERV4 familiak ERV gehien zituztenak izan ziren (9 eta 14 ERVz hurrenez hurren) (III-5. taula). EqERV4 eta EqERV5 familietan izan ezik, non ERVen dibergentzia aldakorra izan zen eta erlazio erradiala azaltzen zuten, talde bakoitzeko ERVen arteko erlazioa politomikoa izan zen eta dibergentzia handia azaldu zuten (2. irudia). Klase honetako familia gehien ordezkariek *pol* geneko RT eskualdeko LPQG motiboa mantentzen zuten, hainbat aldaerekin familia batzuen ordezkarien kasuan. EqERV8 eta EqERV9 familiak, aldiz, sekuentzia nahiko ezberdina azaldu zuten. Era berean, I klaseko familiek YVDD motiboa eta motibo honen aldaera ezberdinak azaldu zituzten; EqERV8 familiaren kasuan, indel baten ondorioz, irakurtaraua aldatua zegoen eta, hortaz, gainontzeko familien ordezkariekin konparatuta, ezberdina izan zen. (III-4. irudi gehigarria). Luzerari dagokionez, I klaseko ERVen luzera-heina 8319 base eta 10837 artekoa izan zen, ERVerik luzeena EqERV3 familiakoa izan zen (11421 base) eta laburrena EqERV9-koa (6695 base). Txertaketa-datari dagokionez (III-5. taula), estimazioa egin ahal izan zen familietan, familia berriena EqERV7 izan zen, familiaren ERVerik zaharrena duela 52 eta 24 miloi urte artean txertatu baitzen. Familiarik zaharrena EqERV4 izan zen,

bere kiderik zaharrena orain dela 208 eta 96 miloi urte artean txertatu baitzen. Gainera, familia honek, hots EqERV4, I klasekoen artean txertatu den ERVrik gazteena azaldu zuen, orain dela 2 eta 1 milioi urte artean txertatutakoa.

III-5. taula: Detektatutako EqERV familiak.

Familiak behintzat bi metodok eta euren *pol* genea >500 nukleotido zuten EqERVekin definitu ziren. Kide bakarreko familien kasuan bakarrik luzeraren media adierazten da. Hainbat familietan txertaketa -data eta ordezkariaren PBSa ezin izan ziren aurrean. Kr, kromosoma.

Detektatutako		Luzera (base)			ERV ordezkaria ^a	Txertaketa-data (ODMU)	
Familia	ERV kopurua	Media	Heina	PBS	Kr:hasi-bukatu (harizpia)	Zaharrena	Gazteena
I Klasea							
EqERV1	9	10676	9922-11288	Tyr	X:53124020-53135308 (+)	71-32	46-21
EqERV2	2	8319	7006-9632	-	7:57912315-57919321 (-)	-	-
EqERV3	6	10837	10028-11421	Tyr	1:27393305-27404726 (+)	77-35	6-3
EqERV4	14	10153	8101-11173	Pri	2:11712828-11722965 (+)	208-96	2-1
EqERV5	7	9714	7525-10351	Ser	10:12944449-12951974 (+)	75-35	30-14
EqERV6	2	8604	7241-9966	-	23:17901636-17911602 (+)	-	-
EqERV7	4	9170	7222-9885	Ser	X:69210501-69217723 (+)	52-24	44-20
EqERV8	2	10007	9918-10095	-	8:3817901-3827996 (-)	97-44	38-18
EqERV9	7	9358	6695-10287	-	20:8848093-8857767 (+)	172-79	41-19
III Klasea							
EqERV10	1	9449		-	29:25492571-25502020 (+)	-	-
EqERV11	9	8990	7188-10983	-	24:27348691-27357972 (+)	179-82	50-23
II Klasea							
EqERV12	3	10577	9879-11956	Lys	5:16125272-16135151 (-)	4-2	1-0.399
EqERV13	2	8336	7503-9168	Lys	13:7964465-7973633 (+)	12-5	12-5
EqERV14	3	9635	9579-9664	Lys	3:25617539-25627203 (-)	40-18	2-1
EqERV15	16	9349	8772-11245	Tyr	1:29475995-29484899 (-)	180-83	10-4

^a Elementurik hurbilena familiaren kontsentsu sekuentziatik.

III.3.2.2. II klasearekin erlazionaturiko familiak. (EqERV12-EqERV15)

Klase honetan EqERV15 familia izan zen ERV kopuru handiena izan zuen taldea (III-5. taula). Klase honetako familien ERVen arteko erlazioa erradiala zen eta, oro har, dibergentzia ez zen oso handia izan (2. irudia). Klase honetako familien ordezkariak, oro har, LPQG motiboa mantendu zuten eta YVDD motiboaren lekuan, aldiz, YMDD motiboa azaldu zuten familia gehienek. (III-4. irudi gehigarria). II klasean, ERVen luzera-heina 8336 eta 10577 base artekoa izan zen. ERVrik luzeena EqERV12 familian aurkitu zen (11956 base) eta laburrena EqERV13-n (7503 base). II klaseko familien artean, EqERV12 familia txertatzen berriena izan zen (III-5. taula), bere ERVrik zaharrenaren txertaketa orain dela 4 eta 2 miloi urte artean gertatu zela estimatu baikenuen eta familiarik zaharrena EqERV15 familia izan zen, 180 eta 83 milioi urte arteko ERV bat familia honetako kiderik zaharrena baitzen. Klase honen barruan, EqERV12 familiak txertaketarik gazteena izan zuen, orain dela 1 eta 0.399 milioi urte artean txertatutako ERVa.

III.3.2.3. III klasearekin erlazionaturiko familiak. (EqERV10-EqERV11)

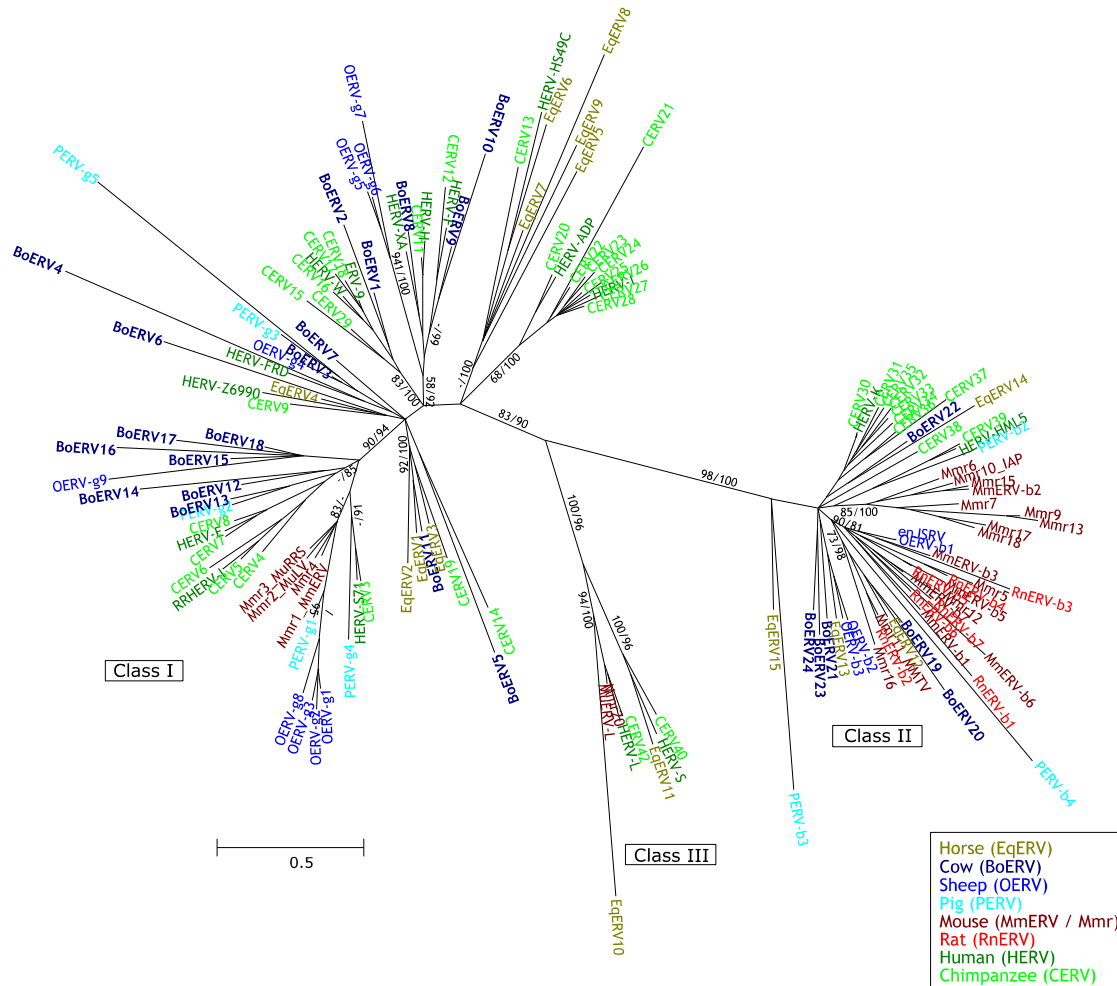
Klase honekin erlazioa zuten ERV gutxi detektatu ziren eta bi familietan sailkatu ziren (III-5. taula). EqERV10 familia elementu bakartia izan zitekeen ERV batez osotua dago, hala ere, familia honek espermentalki lortu ziren zaldiaren ERVekin (AJ233652 - AJ233656) erlazio urruna azaldu zuen. Familia horretako gainontzeko ERVen dibergentzia altua izan zen (III-2. irudia). EqERV10 familiako ordezkariak LPQG motiboa kontserbatua zuen eta YMDD motiboaren YIDD aldaera. EqERV11 familiak, aldiz, LSQG eta YIND motiboak azaldu zituen (III-4. irudi gehigarria). III klaseko familien artean, ERVrik luzeena (10983 base) eta laburrena (7188 base) EqERV11 familian detektatu ziren. III klasean EqERV11 familiaren txertaketa-data estimatzea posible izan zen bakarrik (III-5. taula), txertaketarik zaharrena orain dela 179 eta 82 miloi urte artean gertatu zen eta berriena orain dela 50 eta 23 miloi urte artean.

III.3.3. Espezie ezberdinen arteko erlazio filogenetikoak

Espezie ezberdinetako ERVen zuhaitz filogenetikoan, lan honetan definitutako EqERV familien ordezkariak beste espezieetako ERVekin erlazioa azaldu zuten (III-3. irudia).

Zaldiaren erretrovirus endogenoen (ERV) detekzioa eta sailkapena

Espezie ezberdinetako ERVak argi banatu ziren hiru klaseetan eta, orokorrean, ERVen erlazio filogenetikoa jatorrizko espeziearen erlazio filogenetikoa jarraitu zuten. EqERV5, 6, 7, 8, 9, 10, 11 familiak txinpantze eta gizakien ERVekin batera agertu ziren modu sendoan eta EqERV4 familia gizaki eta txinpantzeen ERVekin agertzen bada ere, topologiak ez zuen sendotasun estatistiko nahikorik izan. EqERV1, 2, 3, 12 familiak behiaren ERVekin batera agertu ziren modu sendoan eta EqERV13, 14, 15 familien kasuan, behiarekin azaldu zuten erlazioak ez ziren esangarriak izan.



III-3. irudia: Espezie ezberdinetako ERVen RT eskualdean oinarritutako errotu gabeko zuhaitz filogenetikoa.

Inferentzia bayestarra, 10^6 belaunaldi. Adarretan ondorengo probabilitate Bayestarra adierazten da.

III.3.4. Behiaren eta zaldiaren ERVen detekzioen konparaketa

Behian egindako detekzioarekin konparatuz (III-6. taula), zaldian erabilitako hiru metodoek ERV gutxiago detektatu zituzten, BLASTek 2,4 aldiz gutxiago, LTR_STRUCek 15,42 aldiz gutxiago eta Retrorector©-ek 6,01 aldiz gutxiago. Elementu hauek, antzeko proportzioetan, genomak duten presentzia txikiagoa izan zuten ere (BLASTen 2,25 aldiz gutxiago; LTR_STRUCen 14,75 aldiz gutxiago; Retrorector©-en 6,40 aldiz gutxiago).

ERVen klaseei dagokiela (III-6. taula), behian ez bezala, BLAST metodoaren bidez zaldian III klaseko ERVak detektatu ziren. Metodo honen bitartez, zaldian I klaseko ERV gutxiago detektatu ziren eta II klasekoak behian duten antzeko proportzioan. Bestalde, Retrorector©-en kasuan, zaldian eta behian detektatutako ERVak antzeko proportzioan klase ezberdinetan sailkatu ziren. Familiei dagokionez, behian baino familia gutxiago detektatu ziren, behian 24 familia detektatu baitziren eta zaldian 15.

ERVen banaketa genomikoa metodoz metodo aldatzen bada ere, banaketa homogeneotik aldendu ziren kromosoma-kopurua antzekoa izan zen. Behian gertatu zen bezala, banaketa orokorra ez zen homogeneoa izan (III-6. taula).

Genoma osoko ERVen bilaketak burutu diren gainontzeko espezieekin konparatuz (III-2. taula gehigarria), zaldian detektatutako ERV kopurua txikiagoa izan zen erabilitako metodo guztietan. Era berean, ERV hauen luzera genoma osoarekiko beste espezieetan baino txikiagoa izan zen. Klasez klaseko banaketan, I klasekoak gailendu ziren, primateetan bezala eta III klasekoak beste espezieetan baino gutxiago. Familia-kopurua ere beste espezieetan baino txikiagoa izan zen. Azkenik, gizakian bezala, ERVen banaketa kromosomikoa ez zen homogeneoa izan (III-2. taula gehigarria).

III-6. taula: Behiaren eta zaldiaren ERVen detekzioaren konparazioa.

Espezia	Behia ^d			Zaldia (EquCab2) ^e		
Familiak	24			15		
Metodoa	BLAST	LTR_STRUC	Retrorector©	BLAST	LTR_STRUC	Retrorector©
Kopurua	928	4487	9698	378	291	1615
Estaldura ^a	0.36	1.77	4.29	0.16	0.12	0.67
Klaseak (%)						
I	81.78		61.67	70.94 ^f		63.42 ^g
II	18.22		12.81	20.51 ^f		11.27 ^g
III	0		1.25	8.55 ^f		2.04 ^g
Sailkatzeke			24.27			22.97 ^g
Banaketa						
Gehiago	2, X	3	5, X	2, X	2, X	4, X
kromosomikoa ^b						
Gutxiago	2	4	7	3	1	7

- Daturik ez.

^a Portzentaia sekuentziatutako genomaren arabera kalkulatu zen.

^b Banaketa homogeneo batekin konparatuta, itxarotakoak baino gehiago edo gutxiago. Autosomiko kopurua adierazi da.

^d II kapitulua. ^e lan hau.

^f III-2. irudian identifikatutako elementuetan oinarritu da kalkulua

^g Retrorector©-en sailkapenean oinarrituta. Sailkatzeke, bi edo hiru klasetan sailkatu ziren ERVen proportzioa.

III.4. Eztabaida

Lan honetan zaldiaren genomaren erretrovirus endogenoen (ERV) detekzioa izan da lehenengo helburua. Hortaz lehen eztabaida puntua detekzioa bera da. Lan honetan behian burututako ERVen detekzioaren [II kapitulua] erkaketarako informazio baliagarria ematen digu.

Behiaren kasuan esan bezala [II kapitulua], lan honetan aurkeztu diren emaitzak benetako ERVen kopuruaren azpiestima izan daitezke. Behiaren kasuan informazio kromosomikoa erabiltzeak azpiestima honen arrazoiak hartu genuen, analizatu gabe gelditzen ziren *contigen* ondorioz. Analisi honetan bi genoma-bertsio analizatzeko aukeran oinarrituta ideia honetan sakondu dezakegu. X kromosoma aipatutakoaren adibide argia da. EquCab1 bertsioan *contig* gutxi batzuk zeuden kokatuta eta EquCab2 bertsioan X kromosoma *contig* gehiagoren informazioarekin hornitu zen. Informazio-handitze hori detekzioan islatzen da eta, ondorioz, detektatutako ERV kopurua nabarmenki handitu zen, BLASTen oinarritutako detekzioan ERVen kopurua %300an emendatuz, LTR_STRUCen %1900an eta Retrorector©-en %477.41an.

Honetaz gain, bilaketa-estrategiak (programak barne) hartu genituen detekzioaren beste mugatzat. Zaldiaren genomaren behiaren genomaren programa bakoitzak zuen errendimenduari buruzko orokortasun batzuk errepikatu ziren eta beste batzuk ez. Programa bakoitzak aurkitu zituen ERVen egitura antzeko emaitzak lortu ziren. Oro har, behian ikusitako orokortasunak zaldiaren bi genoma-bertsioetan errepikatu ziren: metodo bakoitzak detektatzen dituen “elementu oso”-en proportzioa antzekoa izan zen, LTR_STRUCek *pol* gabeko elementu gehiago detektatu zituen, gainontzeko metodoetan *pol*dun elementuak gailentzen ziren bitartean. Zaldian behian baino *pol*dun elementuen proportzioa handiagoa izan zen behiarekin konparatuta. Bestalde, Retrorector©-ek eta LTR_STRUCek detektatutako elementuen egituren proportzioak, bai behian zein zaldiaren bi bertsioetan antzekoak izan ziren.

Aipatutako antzekotasunez gain ezberdintasun batzuk nabarmendu ziren ere. Alde batetik zaldian BLASTen oinarritutako metodoaren bidez LTR_STRUC programarekin baino balizko ERV gehiago detektu ziren. Behian, aldiz, kontrakoa gertatu zen. Bestetik, BLASTen bidez detektatutako ERVen egitura zaldien genoma-bertsio ezberdinetan antzeko proportzioa azaldu bazuten ere, behiarekin alderatuz proportzioak ezberdinak izan ziren.

Behian gertatu zen bezala, ERVen banaketa homogeenoa batetik aldentzen ziren kromosomak ezberdinak izan ziren metodoz metodo. Are gehiago, informazio hau zaldiaren genomaren bertsio batetik bestera aldatu zen. Hortaz, ERVen banaketa homogeenoearen analisia zeharo detekzioaren menpekkoa da, hots, erabilitako metodoa eta eskuragarri dagoen informazioaren menpekkoa. Hala ere, gizakian (Villesen et al., 2004) eta behian [II kapitulua] gertatzen den bezala, zaldiaren X kromosomak banaketa homogeenoa batean baino ERV gehiago edukitzea oso esangarria izan zen. Gertaera honen arrazoia argi ez badago ere, bi ikerketa kontutan hartu beharko genituzke. Alde batetik, gizakiaren X kromosomaren %29a L1 LINE familiak osotuta dago (Ross et al., 2005) eta elementu errepikakor honen presentzia mamaliotan X kromosomaren inaktibazioarekin lotua izan da (Bailey et al., 2000; Ross et al., 2005). Bestetik, gizakiaren ERV familia bat, HERV-W deitutakoa hain zuzen ere, LINEen bidez ugaritu direla jakina da (Belshaw et al., 2005). Hortaz, bi behaketa hauek uztartuz, LINEen eraginez ERVak X kromosoman itxarotakoa baino kopuru altuagoan agertzea posible izan daiteke.

Metodo bakoitzak detektatutako elementu kopuruek kromosomen tamainarekin korrelazioa azaldu zuten baina ez euren GC eduki, gene-dentsitate eta pseudogene-dentsitatearekin. Behiaren analisiarekin alderatuz, behian detektatutako elementu-kopurua eta GC edukiaren arteko korrelazio negatiboaren [II kapitulua] detekzioa izan zen ezberdintasun bakarra.

Detektatutako ERVak hiru klase erretrobiraletan sailkatu ziren. Lehendik, esperimentalki II (Hecht et al., 1996) eta III (Benit et al., 1999) klaseko ERVak ezagunak ziren. I klaseko ERVak, ornodun gehienetan bezala, gehiengo osoz badute ere, hemen lehen aldiz aurkeztzen da euren detekzioa. RT eskualdea kontserbatuta zuten ERVak eta esperimentalki lortu ziren ERVak (Benit et al., 1999) ez zuten erlazio esturik izan. Hala ere, sekuentzia horiek EqERV10 familiarekin erlazioa zutela ikusi da eta, agian, familia horretan barneratu daitezke.

Espezie ezberdinetako ERV ordezkarien zuhaitzean zaldiaren familien ordezkariak bi eratan banatu ziren: gizaki/txinpantzeekin multzokatzen zirenak eta behiarekin multzokatzen zirenak. III klaseko ERVen kasuan behiarekin batera ez agertzea ez da harritzekoa, alde batetik, sekuentzia gutxi eskuragarri zeudelako eta, bestetik, euren jatorria mamaliotan nahiko antzinakoa izan daitekeela aditzera eman delako (Herniou et al., 1998; Benit et al., 1999). I klaseko familien kasuan primate/zaldi erlazio honek jakin-mina sortzen du. Inferentzia bayestarrean erlazio hori guztiz sendoa da baina ML bidezko analisisian talde hau ez da agertu. Datu hauekin ezin daiteke jakin taldekapen hau artefaktu bat den, antzeko ERV biren infekzio independente bat den edo transmisio horizontal bat den. Gainontzeko familien kasuan, behi eta zaldien arteko erlazioa itxarotekoa da, biak Laurasiatheria super ordeneko kideak baitira. Hots, espezie hauek ahaideak diren heinean, ERVen arteko erlazioa aitzinekoa izatea itxaron daiteke.

Zaldian egon daitezkeen ERVak beste espezieekin konparatzerakoan, zaldian ERV gutxi detektatu zirela argi dago; txakurra eta katuan bezain gutxi edo gutxiago. Talde filogenetiko bereko espezieak aztertzerakoan gizaki/txinpantze, sagu/arratoi, txakur/katu euren artean ezberdintasun gutxi ikusi ziren. Hala ere, taldeen artean ezberdintasunak zeuden. Hots, talde filogenetiko bakoitzean gertatzen dena genoma bakoitzaren dinamikaren menpe egon daitekeela dirudi (Jern & Coffin, 2008; Stocking & Kozak, 2008).

Bestalde, espezie batek jaso dituen infekzio erretrobiralen historia eta horiei aurre egiteko gaitasuna euren genomak dute isla, ERVak erregistro "fosila" bilakatuz (Kaiser et al., 2007; Gifford et al., 2008). Adibidez, txinpantzean agertzen den PtERV1/CERV1 familia ez da gizakian agertzen.

ERV hori “berpiztuz”, hots, familia horren arbasoaren sekuentzia berreraikiz eta infektagarri bilakatzuz, txinpantzearen zelulak infektatzeko gaitasuna azaldu zuen. Gizakian, aldiz, infektatzeko gaitasun oso murriztua izan zuen (Kaiser et al., 2007). Ezberdintasun hau ostalariaren defentsa-mekanismoarekin lotu zen, gizakiaren TRIM5 α proteinak ERV horren aurrean txinpantzearenak baino eraginkortasun handiagoa azaldu baitzuen (Kaiser et al., 2007). Hortaz, behar bada, zaldiak beste espeziek baino infekzio erretrobiralekiko esposizioa txikiagoa izan delako edota horien aurrean defentsa-erantzun hobea izan duelako, ERV gutxiago detektatzen dira.

Birusen infekzioen dinamikarekin eta animalien bizimoduarekin lotuta, goriletan ebola birusarekin egindako eredu simulatuetan (Nunn et al., 2008) hainbat parametrok infekzioaren garapenarekin lotura dutela aditzera eman zuten. Ebola transmisio zuzena duen patogeno birulentoa da eta eredu hau goriletan aztertu bazen ere, eta zaldiaren populazioak naturala ez badira ere, egitura sozial finkoa duen edozein populazio ostalariari aplikatu daitekeela iradokitzen dute egileek. Infekzio-hedapenaren eredu honetatik lortu zen ondorio interesgarri bat taldeen egitura sozialean zegoen. Simulazioak ar bat eta hainbat emeez osotutako gorila taldeekin burutu ziren. Talde ezberdinen arteko infekzioa emeen kopuruak baldintzatu zezakeela aditzera eman zen, batez ere taldea sei emez baino gehiagoz osotuta bazegoen. Kasu horietan, infekzioaren ondorioz arra hiltzen bazen, eme horiek beste talde batzuetara joaten ziren eta, ondorioz, infekzioa zabaltzeko probabilitatea handiagoa zegoen. Horrela, sei emez baino gehiagoz osotutako taldeek, infekzioen kasuetan, beste taldeetara infekzioa eramateko arrisku handiagoa zutela ondorioztatu zuten (Nunn et al., 2008). Zaldiek ere halako taldeak osatzen zituzten (Sorin, 2001) eta, honek, eraginaren bat izan dezake erretrobirusen hedapenarekin.

Behiaren eta zaldiaren genomen artean ikusten diren ezberdintasunak ulertzeko beste faktore gako bat etxekotze-prozesua izan daiteke. Bi espezieen etxekotze prozesua ezberdina izan da: behia orain dela 8.000-10.000 urte artean etxekotu zen, bi gertaera independentetan (Loftus et al., 1994) eta zaldia orain dela 2000-4500 urte artean, eta bere etxekotze prozesuan hainbat populazio ezberdinek parte hartu zuten (Jansen et al., 2002). Zaldiaren kasuan, detektatu ziren ERVen artean eta euren txertaketa-data estimatzea posible izan zirenen artean, guztiak etxekotzea gertatu baino lehenagokoak izan ziren, behian ez bezala, non etxekotze-prozesu ostean txertatu diren ERVak detektatu ziren. Hortaz, ezin daiteke baztertu etxekotze prozesu ezberdinek infekzioen dinamikan eragitea eta, ondorioz, ERVak genoma jakin batean txertatzeko probabilitatea. Gainera, etxekotu diren animalien taldeen tamaina kontutan hartu behar da: behiak talde handiagotan aurkitzen baitira eta zaldiak talde txikiagotan. Elementu transposakorren kopurua populazioaren tamainaren eta zorizko prozesuen menpekoea izan daitekeela proposatu denez (Biemont & Vieira, 2006) taldeen tamainen eragina ezin daiteke baztertu.

Laburbilduz, lan honetan zaldiaren genoman lehen ERVen bilaketa sistematikoa aurkeztu da. Antzeko moduan aztertutako genometan baino ERV gutxiago detektatu dira zaldian, honek, genoma bakoitzean edo, behintzat, talde taxonomiko ezberdinetako genometan ERVen txertaketa ezberdina dela adieraz lezake. Hortaz, orain ikusten dugun ERVen kopurua talde taxonomiko bakoitzak jaso dituen infekzio erretrobiralen menpekoea izan daiteke.

III.5. Bibliografia

- Altschul, S., T. Madden, A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D. Lipman. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* **25**:3389-3402.
- Bailey, J., L. Carrel, A. Chakravarti, and E. Eichler. 2000. Molecular evidence for a relationship between LINE-1 elements and X chromosome inactivation: The Lyon repeat hypothesis. *Proceedings of the National Academy of Sciences of the United States of America* **97**:6634-6639.
- Baillie, G., L. van de Lagemaat, C. Baust, and D. Mager. 2004. Multiple groups of endogenous betaretroviruses in mice, rats, and other mammals. *Journal of Virology* **78**:5784-5798.
- Belshaw, R., A. Katzourakis, J. Paces, A. Burt, and M. Tristram. 2005. High copy number in human endogenous retrovirus families is associated with copying mechanisms in addition to reinfection. *Mol Biol Evol* **22**:814-817.
- Benit, L., J. Lallemand, J. Casella, H. Philippe, and T. Heidmann. 1999. ERV-L elements: a family of

endogenous retrovirus-like elements active throughout the evolution of mammals. *Journal of Virology* **73**:3301-3308.

Biemont, C. and C. Vieira. 2006. Genetics: Junk DNA as an evolutionary force. *Nature* **443**:521-524.

Boeke, J. and J. Stoye. 1997. Retrotransposons, endogenous retroviruses, and the evolution of retroelements, p. 343-435. *In* J. Coffin, S. Hughes, and H. Varmos (eds.), *Retroviruses*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Castresana, J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Biol Evol* **17**:540-552.

Gifford, R., P. Kabat, J. Martin, C. Lynch, and M. Tristem. 2005. Evolution and distribution of class II-related endogenous retroviruses. *Journal of Virology* **79**:6478-6486.

Gifford, R., A. Katzourakis, M. Tristem, O. Pybus, M. Winters, and R. Shafer. 2008. A transitional endogenous lentivirus from the genome of a basal primate and implications for lentivirus evolution. *Proceedings of the National Academy of Sciences of the United States of America* **105**:20362-20367.

Griffiths, D. J. 2001. Endogenous retroviruses in the human genome sequence. *Genome Biol* **2**:REVIEWS1017.

Guindon, S. and O. Gascuel. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic Biology* **52**:696-704.

Hecht, S., K. Stedman, J. Carlson, and J. DeMartini. 1996. Distribution of endogenous type B and type D sheep retrovirus sequences in ungulates and other mammals. *Proceedings of the National Academy of Sciences of the United States of America* **93**:3297-3302.

Herniou, E., J. Martin, K. Miller, J. Cook, M. Wilkinson, and M. Tristem. 1998. Retroviral diversity and distribution in vertebrates. *Journal of Virology* **72**:5955-5966.

Huda, A., N. Polavarapu, I. Jordan, and J. McDonald. 2008. Endogenous retroviruses of the chicken genome. *Biology Direct* **3**.

Jansen, T., P. Forster, M. A. Levine, H. Oelke, M. Hurles, C. Renfrew, J. +. Weber, and K. Olek. 2002. Mitochondrial DNA and the origins of the domestic horse. *Proceedings of the National Academy of Sciences of the United States of America* **99**:10905-10910.

Jern, P. and J. Coffin. 2008. Effects of Retroviruses on Host Genome Function. *Annual Review of Genetics* **42**:709-732.

Johnson, W. and J. Coffin. 1999. Constructing primate phylogenies from ancient retrovirus sequences. *Proceedings of the National Academy of Sciences of the United States of America* **96**:10254-10260.

Kaiser, S. M., H. S. Malik, and M. Emerman. 2007. Restriction of an Extinct Retrovirus by the Human TRIM5(Kaiser et al., 2007) Antiviral Protein. *Science* **316**:1756-1758.

Katoh, K., K. Kuma, H. Toh, and T. Miyata. 2005. MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Research* **33**:511-518.

Katzourakis, A., M. Tristem, O. Pybus, and R. Gifford. 2007. Discovery and analysis of the first endogenous lentivirus. *Proceedings of the National Academy of Sciences of the United States of America* **104**:6261-6265.

Klymiuk, N., M. Muller, G. Brem, and B. Aigner. 2003. Characterization of endogenous retroviruses in sheep. *Journal of Virology* **77**:11268-11273.

Kumar, S., K. Tamura, and M. Nei. 2004. MEGA3: Integrated software for molecular evolutionary genetics analysis and sequence alignment. *Briefings in Bioinformatics* **5**:150-163.

Loftus, R. T., D. E. MacHugh, D. G. Bradley, P. M. Sharp, and P. Cunningham. 1994. Evidence for two independent domestications of cattle. *Proceedings of the National Academy of Sciences of the United States of America* **91**:2757-2761.

McCarthy, E. and J. McDonald. 2003. LTR_STRUC: a novel search and identification program for LTR retrotransposons. *Bioinformatics* **19**:362-367.

McCarthy, E. and J. McDonald. 2004. Long terminal repeat retrotransposons of *Mus musculus*. *Genome Biology* **5**.

Nunn, C., P. Thrall, K. Stewart, and A. Harcourt. 2008. Emerging infectious diseases and animal social systems. *Evolutionary Ecology* **22**:519-543.

Patience, C., W. Switzer, Y. Takeuchi, D. Griffiths, M. Goward, W. Heneine, J. Stoye, and R. Weiss. 2001. Multiple groups of novel retroviral genomes in pigs and related species. *Journal of Virology* **75**:2771-2775.

Polavarapu, N., N. Bowen, and J. McDonald. 2006. Identification, characterization and comparative genomics of chimpanzee endogenous retroviruses. *Genome Biology* **7**.

Posada, D. and K. Crandall. 1998. MODELTEST: testing the model of DNA substitution. *Bioinformatics* **14**:817-818.

R Development Core Team. 2008. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Rice, P., I. Longden, and A. Bleasby. 2000. EMBOSS: The European molecular biology open software suite. *Trends in Genetics* **16**:276-277.

Ronquist, F. and J. Huelsenbeck. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models.

Bioinformatics 19:1572-1574.

Ross, M., D. Grafham, A. Coffey, S. Scherer, K. McLay, D. Muzny, M. Platzer, G. Howell, C. Burrows, C. Bird, A. Frankish, F. Lovell, K. Howe, J. Ashurst, R. Fulton, R. Sudbrak, G. Wen, M. Jones, M. Hurles, T. Andrews, C. Scott, S. Searle, J. Ramser, A. Whittaker, R. Deadman, N. Carter, S. Hunt, R. Chen, A. Cree, P. Gunaratne, P. Havlak, A. Hodgson, M. Metzker, S. Richards, G. Scott, D. Steffen, E. Sodergren, D. Wheeler, K. Worley, R. Ainscough, K. Ambrose, M. Ansari-Lari, S. Aradhya, R. Ashwell, A. Babbage, C. Bagguley, A. Ballabio, R. Banerjee, G. Barker, K. Barlow, I. Barrett, K. Bates, D. Beare, H. Beasley, O. Beasley, A. Beck, G. Bethel, K. Blechschmidt, N. Brady, S. Bray-Allen, A. Bridgeman, A. Brown, M. Brown, D. Bonnin, E. Bruford, C. Buhay, P. Burch, D. Burford, J. Burgess, W. Burrill, J. Burton, J. Bye, C. Carder, L. Carrel, J. Chako, J. Chapman, D. Chavez, E. Chen, G. Chen, Y. Chen, Z. Chen, C. Chinault, A. Ciccodicola, S. Clark, G. Clarke, C. Clee, S. Clegg, K. Clerc-Blankenburg, K. Clifford, V. Copley, C. Cole, J. Conquer, N. Corby, R. Connor, R. David, J. Davies, C. Davis, J. Davis, O. Delgado, D. DeShazo, P. Dhami, Y. Ding, H. Dinh, S. Dodsworth, H. Draper, S. Dugan-Rocha, A. Dunham, M. Dunn, K. Durbin, I. Dutta, T. Eades, M. Ellwood, A. Emery-Cohen, H. Errington, K. Evans, L. Faulkner, F. Francis, J. Frankland, A. Fraser, P. Galgoczy, J. Gilbert, R. Gill, G. Glockner, S. Gregory, S. Gribble, C. Griffiths, R. Grocock, Y. Gu, R. Gwilliam, C. Hamilton, E. Hart, A. Hawes, P. Heath, K. Heitmann, S. Hennig, J. Hernandez, B. Hinzmann, S. Ho, M. Hoffs, P. Howden, E. Huckle, J. Hume, P. Hunt, A. Hunt, J. Isherwood, L. Jacob, D. Johnson, S. Jones, P. de Jong, S. Joseph, S. Keenan, S. Kelly, J. Kershaw, Z. Khan, P. Kioschis, S. Klages, A. Knights, A. Kosiura, C. Kovar-Smith, G. Laird, C. Langford, S. Lawlor, M. Leversha, L. Lewis, W. Liu, C. Lloyd, D. Lloyd, H. Loulseged, J. Loveland, J. Lovell, R. Lozado, J. Lu, R. Lyne, J. Ma, M. Maheshwari, L. Matthews, J. McDowall, S. McLaren, A. McMurray, P. Meidl, T. Meitinger, S. Milne, G. Miner, S. Mistry, M. Morgan, S. Morris, I. Muller, J. Mullikin, N. Nguyen, G. Nordsiek, G. Nyakatura, C. O'Dell, G. Okwuonu, S. Palmer, R. Pandian, D. Parker, J. Parrish, S. Pasternak, D. Patel, A. Pearce, D. Pearson, S. Pelan, L. Perez, K. Porter, Y. Ramsey, K. Reichwald, S. Rhodes, K. Ridler, D. Schlessinger, M. Schueler, H. Sehra, C. Shaw-Smith, H. Shen, E. Sheridan, R. Shownkeen, C. Skuce, M. Smith, E. Sotheran, H. Steingruber, C. Steward, R. Storey, R. Swann, D. Swarbreck, P. Tabor, S. Taudien, T. Taylor, B. Teague, K. Thomas, A. Thorpe, K. Timms, A. Tracey, S. Trevanion, A. Tromans, M. d'Urso, D. Verduzco, D. Villasana, L. Waldron, M. Wall, Q. Wang, J. Warren, G. Warry, X. Wei, A. West, S. Whitehead, M. Whiteley, J. Wilkinson, D. Willey, G. Williams, L. Williams, A. Williamson, H. Williamson, L. Wilming, R. Woodmansey, P. Wray, and J. Yen. 2005. The DNA sequence of the human X chromosome. *Nature* 434:325-337.

Smit, AFA, Hubley, R, and Green, P. RepeatMasker Open-3.0. 2004. Ref Type: Internet Communication

Sorin, A. "Equus caballus" (On-line). Animal Diversity Web . 2001. Ref Type: Electronic Citation

Sperber, G., T. Airola, P. Jern, and J. Blomberg. 2007. Automated recognition of retroviral sequences in genomic data - RetroTector (c). *Nucleic Acids Research* 35:4964-4976.

Stephens, R., J. Casey, and N. Rice. 1986. Equine Infectious-Anemia Virus gag and pol genes - relatedness to Visna and AIDS virus. *Science* 231:589-594.

Stocking, C. and C. Kozak. 2008. Murine endogenous retroviruses. *Cellular and Molecular Life Sciences* 65:3383-3398.

Thompson, J., D. Higgins, and T. Gibson. 1994. Clustal-W - Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research* 22:4673-4680.

Tobaly-Tapiero, J., P. Bittoun, M. Neves, M. C. Guillemin, C. H. Lecellier, F. Puvion-Dutilleul, B. Gicquel, S. Zientara, M. L. Giron, H. de The, and A. Saib. 2000. Isolation and Characterization of an Equine Foamy Virus. *J. Virol.* 74:4064-4073.

Tristem, M. 2000. Identification and characterization of novel human endogenous retrovirus families by phylogenetic screening of the Human Genome Mapping Project database. *Journal of Virology* 74:3715-3730.

Villesen, P., L. Aagaard, C. Wiuf, and F. S. Pedersen. 2004. Identification of endogenous retroviral reading frames in the human genome. *Retrovirology* 1.

Wade, C. M., E. Giulotto, S. Sigurdsson, M. Zoli, S. Gnerre, F. Imsland, T. L. Lear, D. L. Adelson, E. Bailey, R. R. Bellone, H. Blocker, O. Distl, R. C. Edgar, M. Garber, T. Leeb, E. Mauceli, J. N. MacLeod, M. C. T. Penedo, J. M. Raison, T. Sharpe, J. Vogel, L. Andersson, D. F. Antczak, T. Biagi, M. M. Binns, B. P. Chowdhary, S. J. Coleman, G. la Valle, S. Fryc, G. Guerin, T. Hasegawa, E. W. Hill, J. Jurka, A. Kiialainen, G. Lindgren, J. Liu, E. Magnani, J. R. Mickelson, J. Murray, S. G. Nergadze, R. Onofrio, S. Pedroni, M. F. Piras, T. Raudsepp, M. Rocchi, K. H. Roed, O. A. Ryder, S. Searle, L. Skow, J. E. Swinburne, A. C. Syvanen, T. Tozaki, S. J. Valberg, M. Vaudin, J. R. White, M. C. Zody, Broad Institute, Broad Institute Whole Genome Assembly Team, E. S. Lander, and K. Lindblad-Toh. 2009. Genome Sequence, Comparative Analysis, and Population Genetics of the Domestic Horse. *Science* 326:865-867.

III.6. Eranskina

III-1. taula gehigarria: Metodo bakoitzak detektatutako elementuen egitura EquCab1 genomaren bertsioan.

Egitura	Retrotector©	LTR_STRUC	BLAST	Guztira ^a
LTR - RT - LTR	646 (%54,29)	29 (%14,43)	-	655 (45,45%)
GAG - POL - ENV	911 (%76,55)		41 (%13,40)	
GAG - POL	224 (%18,82)		57 (%18,63)	
GAG - ENV	16 (%1,35)		29 (%9,48)	
POL - ENV	31 (%2,61)		179 (%58,60)	
GAG	5 (%0,42)			
POL	2 (%0,17)			
Besteak	1 (%0,08)			
<i>pol</i> presentzia	1168 (%98,15)	29 (%14,43)	277 (%90,52)	
<i>pol</i> absentsia	22 (%1,85)	172 (%85,57)	29 (%9,48)	
Guztira	1190	201	306	1441

^a ERV ez erredundanteak.

III-2. taula gehigarria: Espezie ezberdinetako ERVen banaketa eta sailkapena

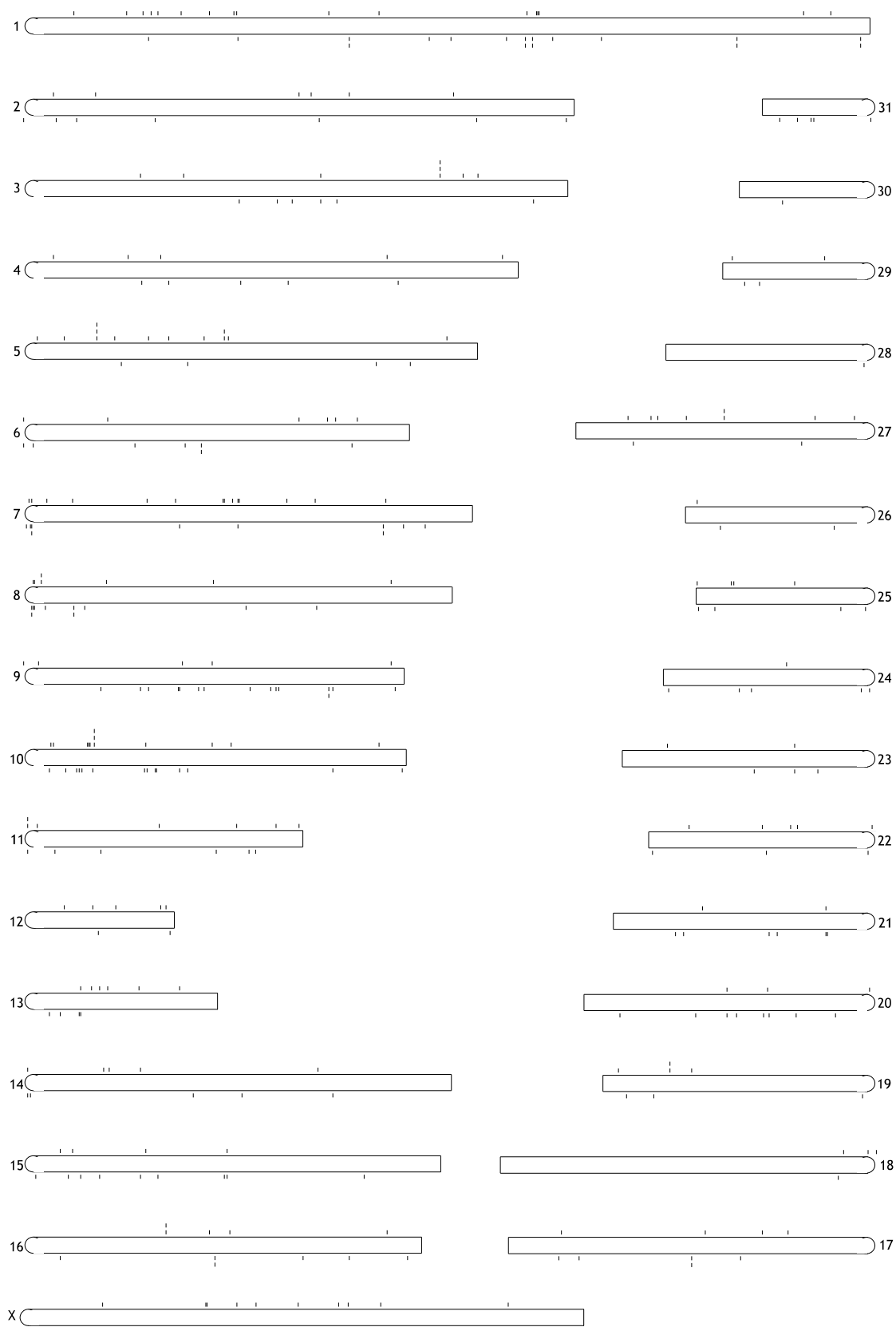
	Kopurua ^a	Estaldura (%) ^a	Klaseak (%)			Families	Banaketa kromosomikoa ^b
			I	II	III		
Gizakia (Repeatmasker)	203000 ^c	4.64 ^c	62.28 ^c	6.68 ^c	31.03 ^c	Klase guztiak 350 ^d / 26 ^{ef}	Gutxiago 9 ^g Gehiago 3, Y ^g
Txinpantzea ^h (LTR_STRUC)	425 oso	-	-	-	-	Klase guztiak 42	-
Makakoa ⁱ (Repeatmasker)	919	-	-	-	-	Klase guztiak 8	-
Sagua (Repeatmasker / BLAST)	-	4.4 ^j	15.46 ^j	71.13 ^j	13.18 ^j	Klase guztiak 20 ^k / II klasea 7 ^l	-
Arratoia (Repeatmasker / BLAST)	-	5.05 ^m	19.21 ^m	64.16 ^m	16.63 ^m	II klasea 7 ^l	-
Txakurra (Repeatmasker) ⁿ	-	1.80	33.8	0	66.12	-	-
Katua (Repeatmasker) ^o	-	2.3	45.6	1.74	52.61	-	-

- Daturik ez.

^a Posible denean ERV kopurua azaltzen da, bestela genomaren estalduraren portzentaia erakusten da. Portzentaia sekuentziatutako genomaren arabera kalkulatu zen.

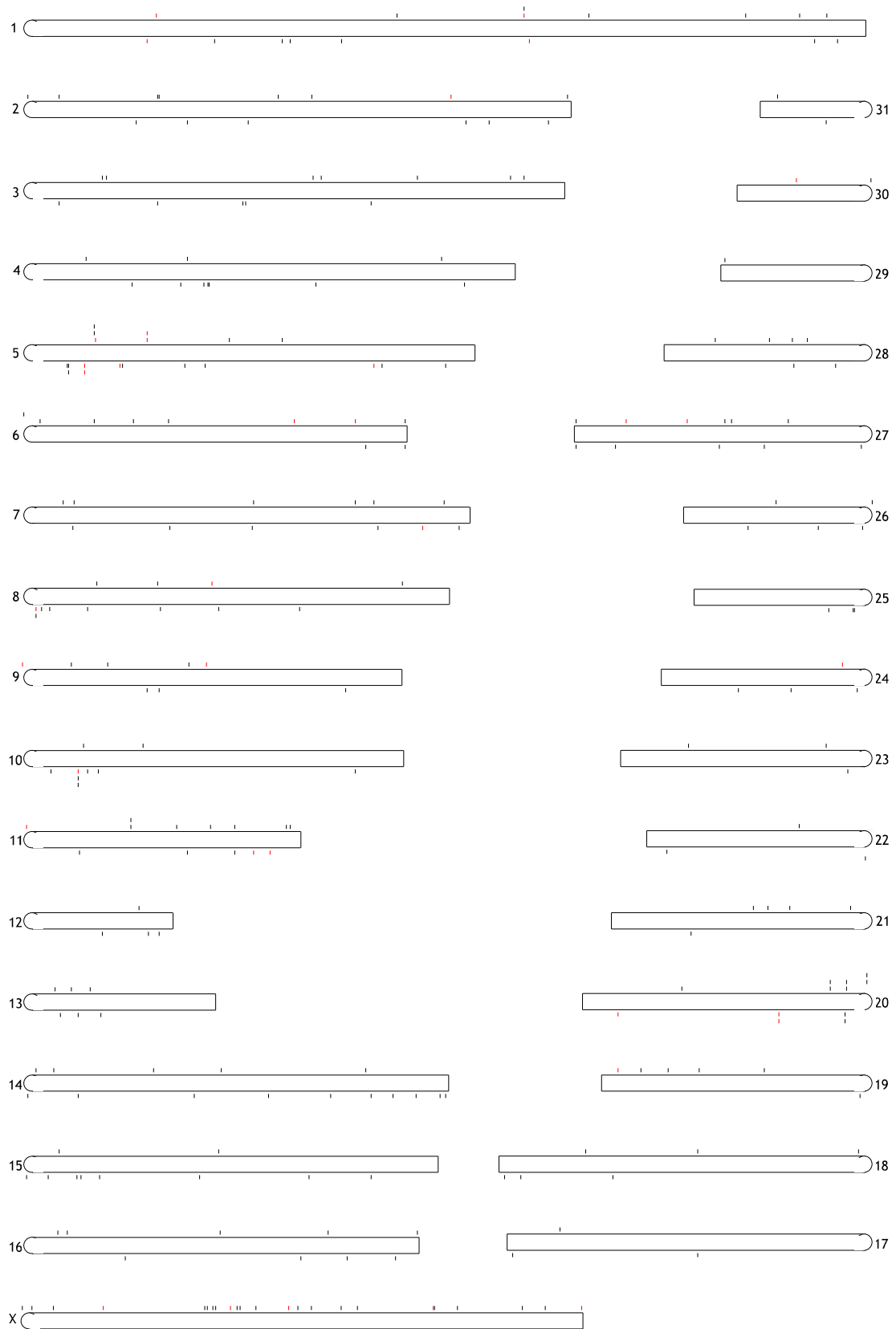
^b Banaketa homogeen batekin konparatuta, itxarotakoak baino gehiago edo gutxiago. Autosomikoen kourua adierazi da.

^c International Human Genome Sequencing Consortium, 2001. ^d Paces et al., 2002. ^e Bénit et al., 2001. ^f Tristem, 2000. ^g Villesen et al., 2004. ^h Polavarapu et al., 2006. ⁱ Han et al., 2007. ^k McCarthy et al., 2004. ^l Baillie et al., 2004. ^m Rat Genome Sequencing Project Consortium, 2004. ⁿ Lindblad-Toh et al., 2005. ^o Pontius et al., 2007.

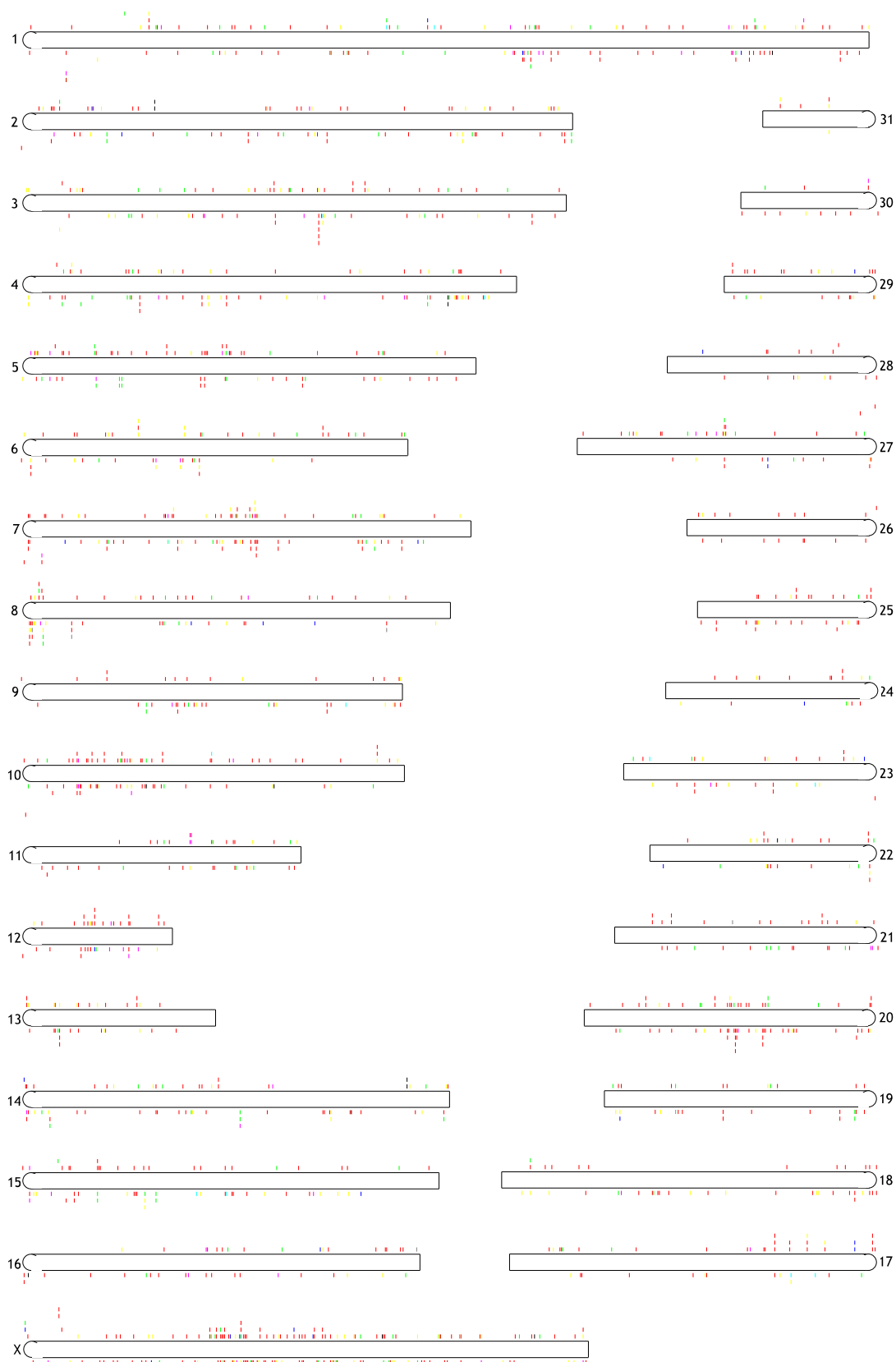


III-1. irudi gehigarria: BLAST bidez detektatutako ERVen kokapen kromosomikoa. Kromosomen gainean (-) harizpia, azpian (+) harizpia.

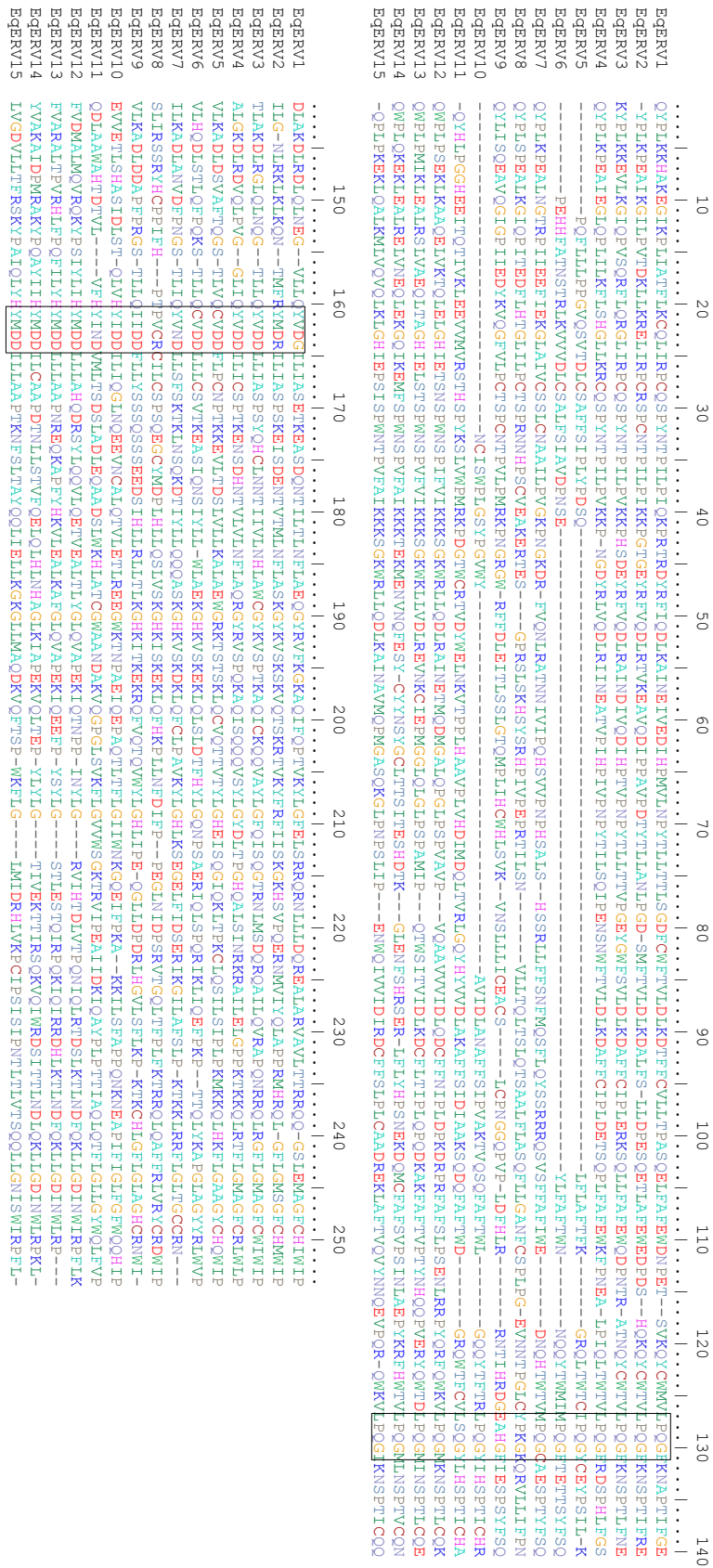
Zaldiaren erretrovirus endogenoen (ERV) detekzioa eta sailkapena



III-2. irudi gehigarria: LTR_STRUC bidez detektatutako LTRdun elementuak. Kromosomen gainean (-) harizpia, azpian (+) harizpia. Gorri RT eskualdea detektatu dien elementuak.



III-3. irudi gehigarria: Retrotector© bidez detektatutako ERVak. Kromosomen gainean (-) harizpia, azpian (+) harizpia. Koloreetan Retrotector©-ek ERV bakoitzari egokitutako klasea: gorria, I klasea; berdea, II klasea, urdina III klasea; horia klase I edo II; magenta, klase I edo III; zian klase II edo III; grisa, zahaztugabe.



III-4. irudi gehigarria: EgERV familien ordezkarien RT eskualdearen lerrokatze partziala.

IV. Mamalioen erretrobirus endogenoen (ERV) dinamika ebolutiboa

IV.1. Sarrera

Erretrobirus endogenoak (ERV, *endogenous retrovirus*) ostalariaren hozi-lerroan txertatu diren erretrobirus exogenoen fase probiralak dira eta belaunaldietan zehar genoma ostalarian txertatuta mantendu dira. ERVak mamalio guztietan eta ornodunen hein zabal batean detektatuak izan dira eta euren egitura tipikoa hiru gene nagusiak (*gag*, *pol* eta *env*) dituen zati zentral batez eta albo bakoitzeko LTRetaz (*Long terminal repeats*) osatuta dago. Erretrobirus exogenoen generoekin duten erlazio filogenetikoaren arabera ERVak, orokorrean, 3 klasetan sailkatzen dira: I klasea, *Gammaretrovirus* eta *Epsilonretrovirus* generoekin erlazioa dutenak; II klasea, *Alpharetrovirus*, *Betaretrovirus*, *Deltaretrovirus* eta *Lentivirus* generoekin erlazionatutakoak; eta III klasea, *Spumavirusekin* erlazioa dutenak (Boeke & Stoye, 1997; Gifford et al., 2005).

Euren antzekotasun filogenetikoaren arabera, genoma bateko ERVak familian multzokatzen dira, hau da, erretrobirus exogeno baten jatorrizko infekzioetik eratorritako kopiak izango liratekeen ERV multzoak. Hortaz, ERV familia hauek ostalariaren genomak kopuru aldakorrean agertzen dira, familia batzuk besteak baino ugariagoak izanik. Behin ostalariaren genomak txertatu direla, ERVak euren kopurua handitzeko gai dira. ERVek bi mekanismo nagusi erabiltzen dituzte euren buruaren kopiak egiteko eta genoma ostalarian ugaritzeko: erretrotransposizioa eta berrinfekzioa (Gifford & Tristem, 2003; Belshaw et al., 2004; Katzourakis et al., 2005).

Erretrotransposizioa deritzon mekanismoan, ERV batek, hozi-lerroko zelulatik irten gabe, bere burua erreplikatzeko du eta zelula horren genomak txertatzen da berrirori. Mekanismo hau bi modutara gerta daiteke: *cis* eran, zeinean ERVak berak kodetutako geneak erabiltzen dituen bere burua kopiatzeko eta txertatzeko; eta *trans* eran, zeinean ERVak beste erretrobirus baten proteinak erabiltzen dituen. Berrinfekzioari dagokionez, ostalari berdinean, ERV batek bere burua kopiatzen du eta hozi-lerroko zelula batetik edo zelula somatiko batetik ateratzen da. Ondoren, hozi-lerroko zelula bat infektatzen du eta zelula honen genomak txertatzen da (Gifford & Tristem, 2003; Belshaw et al., 2004; Katzourakis et al., 2005).

ERVak bere burua erreplikatzeko erabiltzen duen mekanismoa ERVaren kontserbazio mailaren menpekoa da. Erretrotransposizioan, ERVa zelulatik ateratzen ez denez, *env* geneak ez du funtzionala izan behar. (i) *cis* eraren kasuan, ERVak bere proteinak erabiltzen dituen heinean, *gag* eta *pol* funtzionalak izatea beharrezkoa du. (ii) *trans* eran, aldiz, gene nagusiek ez dute funtzionalak izan behar, ERVak beste elementu batzuetatik lortzen baititu. Soilik promotorea eta eskualde gutxi batzuk (polipurina traktua eta hasle-loturaren gunea, adibidez) funtzionalak mantendu behar ditu. Azkenik, (iii) berrinfekzioan, ERVa zelulatik atera behar denez eta beste zelula bat infektatzen duenez, hiru gene erretrobiral nagusiak, hots, *gag*, *pol* eta *env*, funtzional mantendu behar ditu (Gifford & Tristem, 2003; Belshaw et al., 2004; Belshaw et al., 2005; Katzourakis et al., 2005).

Hortaz, kopia-kopuruaren emendioa geneen funtzionaltasunaren arabera aztertzen den heinean, gene erretrobiral nagusien hautespena aztertzea beharrezkoa da. Behin ERVak genoma ostalarian txertatzen direla, gainontzeko material genomikoak jasaten dituen indar ebolutiboen menpe gelditzen dira (Jern & Coffin, 2008). Gainera, ERVen jatorri patogenikoa dela eta, mutazioak pilatzeko eta hautespena erlaxatzeko joera egon daitekela proposatu da (Gifford & Tristem, 2003). Hala ere, gorago aipatu bezala, hainbat ERVek erreplikatzeko gaitasuna mantentzeko gai dira, hots, euren geneek funtzionaltasuna mantentzeko hautespena dago. Mutazioen eragin txikiena jasotzen duen gene erretrobirala *pol* da, eta honen barruan alderantzizko transkriptasa (RT) eskualdea, ERVen erreplikaziorako gakoa den entzima hain zuzen ere (Gifford & Tristem, 2003). Bestalde, *env* geneko TM eskualdeak ere kontserbazio-maila nahiko altua du erretrobirus ezberdinetan (Benit et al., 2001). Beraz, gene erretrobiralen bi eskualde hauek hautespen purifikatzailea jasan ohi dute.

Gizakiaren ERVetan egindako analisisetan, gizakiaren genomak euren kopia-kopurua emendatzeko mekanismo nagusia berrinfekzioa izan dela aditzera eman da (Belshaw et al., 2004; Belshaw et al., 2005). Gainera, azertu ziren ERVen zuhaitz filogenetikoa eraikitzerakoan, barneko adarrek

bukaerako adarrek baino hautespen maila altuagoa azaldu zuten, iraganean aktiboak ziren eta finkatuta ez zeuden ERVek jasandako hautespen presioarekin erlazionatu zena (Belshaw et al., 2004). Hala ere, kopia-kopuru handienak dituzten familiak transposizio bidez erreplikatu ziren, eta badirudi *trans* eraren bidez gertatu zela emendio hau (Belshaw et al., 2005). Gizakiaren ERVetaz gain, beste genoma batzuetan elementu hauen dinamika ez da oso ezaguna.

Erretrobirus exogenoetan, bai gene erretrobiral ezberdinen arteko errekonbinazioa zein gene barruko errekonbinazioa detektatu da. Mekanismo hau aldakortasunaren iturri izan daitekeela eta funtzioak berreskuratzeko baliagarria izan daitekeela proposatu da (Robertson et al., 1995; Vogt, 1997). Euren kide endogenoetan, ordea, mekanismo hau gutxi aztertu bada ere, hainbat adibide aurki ditzakegu. Adibidez, txerriaren ERVen *env* genearen kasuan (Lee et al., 2008), *env* genearen aldaerek euren artean naturalki errekonbinatzen dutela ikusi da, eta gizakiaren ERV familia baten (HERV-V familian, alegia) *env* geneak konbertsio genikoa jasan duela ondorioztatu da (Kjeldbjerg et al., 2008). Espezie hauen ERV familia batzuetan ere *pol* genea aztertu da (Belshaw et al., 2004; Klymiuk et al., 2006). Lan hauetan ERV gutxi batzuen sekuentzia-errekonbinazioa detektatu zen edota hainbat ERVen sekuentzia errekonbinazioz sortu zirela ondorioztatu zen.

Behiaren [II kapitulua] eta zaldiaren [III kapitulua] genometan ERVak detektatu ditugula, egokia dirudi ERV hauen dinamika ebolutiboa aztertzea. Horretarako gene erretrobiral ezberdinen errekonbinazio intergenikoa, errekonbinazio intragenikoa eta hautespena erabiliko dira ERVen kopia-kopuruaren emendioa eta kontserbazioa aztertzeko. Gainera, beste genoma batzuetan gertatzen dena ezagutzeko, ondo sekuentziatutako beste hainbat mamalioetan dauden ERVen dinamika ebolutiboa ere aztertuko da.

IV.2. Material eta metodoak

Behiaren eta zaldiaren genomaz gain, sekuentziazio-estaldura altua duten beste 5 mamalioen genomak aztertu ziren: gizakia, txinpantzea, sagua, arratoia eta txakurra (IV-1. taula).

IV-1. taula: Aztertutako genomak laburpena.

Espezia	Bertsioa	Estaldura	Helbidea
Gizakia (<i>Homo sapiens</i>)	NCBI 36	10X	ftp://ftp.ensembl.org/pub/current_fasta/homo_sapiens/dna/
Txinpantzea (<i>Pan troglodytes</i>)	Pan_troglodytes-2.1	6X	ftp://ftp.ensembl.org/pub/current_fasta/pan_troglodytes/dna/
Sagua (<i>Mus musculus</i>)	NCBI m37	7X	ftp://ftp.ensembl.org/pub/current_fasta/mus_musculus/dna/
Arratoia (<i>Rattus norvegicus</i>)	RGSC 3.4	5X	ftp://ftp.ensembl.org/pub/current_fasta/rattus_norvegicus/dna/
Txakurra (<i>Canis familiaris</i>)	CanFam 2.0	7,6X	ftp://ftp.ensembl.org/pub/current_fasta/canis_familiaris/dna/
Behia (<i>Bos taurus</i>)	Btau_3.1	7,2X	http://www.hgsc.bcm.tmc.edu/projects/bovine/index.html
Zaldia (<i>Equus caballus</i>)	EquCab2	6,8X	ftp://ftp.ensembl.org/pub/current_fasta/equus caballus/dna/

IV.2.1. ERVen detekzioa eta aukeraketa

Aipatutako mamalioen genoma guztietatik erretrobirus endogenoak (ERV) lortzeko amu-sekuentzia gisa ondo karakterizatutako 12 erretrobirus exogenoen *gag* eta *env* geneetako sekuentzia aminoazidikoak erabili ziren (IV-2. taula). *pol* genearen kasuan erretrobirus klase bakoitzeko sekuentzia bat erabili zen, MLV (Murine Leukemia Virus), MPMV (Mason-Pfizer monkey virus) (Baillie et al., 2004) eta BFV (Bovine Foamy Virus) erretrobirusetatik hain zuzen ere. Erabilitako amu-sekuentziak, *pol* genearen kasuan RT (alderantzizko transkriptasa) eskualdearenak izan ziren eta *env* genearen kasuan TM (transmintza) eskualdearenak. Mamalioen genomatan gene erretrobiralen sekuentziak NCBI-BLAST 2.2.14 programako (Altschul et al., 1997) tBLASTn metodoarekin bilatu ziren. Bilaketa hauen emaitzak PHPn idatzitako script-en bidez prozesatuak izan ziren.

ERV posibleak eraikitzeke gene bakoitzaren emaitzak lotu ziren. Bilaketarako erabilitako 12 erretrobirusetan oinarrituta amu-sekuentzien distantziak kalkulatu ziren. Distantzia horien muga

barruan, gutxienez, bi gene erretrobiral agertzen baziren, eskualde hori ERV moduan definitu zen. Modu honetan detektatutako ERVetatik *gag*, *pol* eta *env* geneak zituzten elementuak aukeratu ziren analisiak egiteko, beti ere *pol* genea 500 nukleotido baino luzeagoa bazen.

Hautatutako ERVetan LTRak detektatzeko saiakera egin zen. Horretarako sekuentziak 5000bp urgora eta urbehera luzatu ziren eta LTR_STRUC programa (McCarthy & McDonald, 2003) erabili zen sekuentzia horietan zehar LTRen bikoteak aurkitzeko.

IV-2. taula: Bilaketa genomikoa burutzeko erabilitako erretrobirusak, amu-sekuentzian posizioa eta sekuentzien arteko distantzia

Izena	Generoa	Ostalaria	Accession Number	Amu-sekuentziaren posizioa			Amu-sekuentzien arteko distantziak			
				<i>gag</i>	<i>pol</i>	<i>env</i>	hasiera- <i>gag</i>	<i>gag-pol</i>	<i>pol-env</i>	<i>env-bukatu</i>
I klasea										
GALV	<i>Gammaretrovirus</i>	<i>Hylobates</i>	NC_001885	1255-1888		6998-7555	1255	846	3508	533
MLV	<i>Gammaretrovirus</i>	Muridae	NC_001501	1020-1653	2520-3276	6920-7457	1020	867	3644	875
FeLV	<i>Gammaretrovirus</i>	Felidae	NC_001940	1516-2149		7323-7860	1516	822	3593	588
II klasea										
JSRV	<i>Betaretrovirus</i>	<i>Ovis aries</i>	NC_001494	1347-1667		6502-7111	1347	1444	2830	351
MMTV	<i>Betaretrovirus</i>	Muridae	NC_001503	1181-1774		6786-7362	1181	1194	3077	1443
BLV	<i>Deltaretrovirus</i>	<i>Bos taurus</i>	NC_001414	790-1362		5760-6158	790	572	2874	2261
HTLV	<i>Deltaretrovirus</i>	<i>Homo sapiens</i>	NC_001436	889-1455		5765-6269	889	780	2786	2238
EIAV	<i>Lentivirus</i>	<i>Equus caballus</i>	NC_001450	945-1562		7235-7867	945	807	4107	492
HIV	<i>Lentivirus</i>	<i>Homo sapiens</i>	NC_001802	1035-1401		7727-8261	1035	765	4805	920
Visna	<i>Lentivirus</i>	<i>Ovis aries</i>	NC_001452	973-1566		8176-8806	973	620	5237	396
III klasea										
HSRV	<i>Spumavirus</i>	<i>Homo sapiens</i>	NC_001795	2111-2750		8788-9399	2111	868	4456	2555
BFV	<i>Spumavirus</i>	<i>Bos taurus</i>	NC_001831	2016-2652	3487-4204	8903-9515	2016	835	4699	2487
Bataz beste $\pm$ desbideratze estandar							1373 $\pm$ 564	868 $\pm$ 217	3851 $\pm$ 799	1410 $\pm$ 962

Espezie bakoitzeko ERVak izendatzeko ondoko datuak erabili ziren: zein kromosometan detektatu zen eta kromosoman zuen ordena erabili ziren. Kokapenak IV-1. taula gehigarrian kontsulta daitezke.

IV.2.2. Analisisien planteamendua

Aukeratutako ERVak aztertzeko, espeziez espezie eta gene bakoitzean (*gag*, *pol* eta *env*) aukeratutako sekuentziak eta, taldez-kanpoko bezala, *Drosophilaren* ZAM elementua (AJ000387) lerrokatu ziren MAFFT programaren (Katoh et al., 2005) (linsi aukera) bitartez eta Gblocks programa (Castresana, 2000) erabilita kontserbatutako posizioak lortu ziren (programa honetan erabilitako aukerak: blokearen luzera minimoa: 5; baimendutako hutsuneak: erdiarekin; alboko posiziontzat sekuentzien kopuru minimoa: datu-multzo bakoitzaren erdia; ez-kontserbatutako aldameneko posizioen kopuru maximoa: 10). Gizakiaren, saguaren, txinpantzearen eta arratoiaren kasuan, lerroketa honetan %95ko identitatea baino handiago zuten ERVak ezabatu ziren eta datu multzoan oso antzekoak ziren ERVetatik bakarra utzi zen.

Hautatutako ERVen erlazio filogenetikoak aztertzeko 3 gene erretrobiral nagusiak (*gag+pol+env*) batera erabili ziren. Horretarako MrBayes 3.1 programa (Ronquist & Huelsenbeck, 2003) erabili zen (10^6 belaunaldi eta eredu ebolutiboa finkatu gabe izan ziren erabilitako aukerak).

Zuhaitz filogenetiko honetan oinarrituta ERVen familia ezberdinen multzoak bildu ziren errekonbinazio intrageniko eta hautespenaren analisisetan sakontzeko asmoz. Familien multzo hauei metafamilia deitu diegu. Metafamiliak eraikiz, alde batetik, espezie ezberdinen arteko erkaketa erraztu nahi da, espezie bakoitzean familiak modu ezberdinean definitu baitira; bestetik, errekonbinazioa antzeko ERVen artean gertatzea posible delako (Gifford & Tristem, 2003) eta metafamilietan erlazio filogenetiko sendoa zuten ERVak multzokatu ziren, familia ezberdinekoak baziren ere.

IV.2.3. ERVen kopia-kopuruaren emendioaren analisisa

ERVen kopia-kopuruaren emendioa aztertzeko errekonbinazio intergenikoa eta zuhaitzean zehar barneko eta kanpoko adarrek zuten hautespena erabili ziren. Errekonbinazio intergenikoari esker ERV baten gene guztiek jatorri berdina duten jakin daiteke eta zuhaitzean zeharreko hautespenak aitzinean ERVek jasandako hautespen-presioari buruzko informazioa eskaintzen du.

IV.2.3.1. Errekonbinazio intergenikoa

Errekonbinazioa intergenikoa gertatu den ikusteko gene erretrobiralen analisi filogenetikoa erabili zen. Gene bakoitzaz osotuta eta 3 geneak batera hartuta kontserbatutako posizioez osotutako sekuentziekin 4 zuhaitz filogenetiko eraiki ziren inferentzia bayestarraren bidez: *gag*, *pol*, *env* eta *gag+pol+env* zuhaitz filogenetikoak. Honetarako MrBayes 3.1 programa (Ronquist & Huelsenbeck, 2003) erabili zen (10^6 belaunaldi eta eredu ebolutiboa finkatu gabe izan ziren erabilitako aukerak).

Parte zentraleko geneen analisisiez gain, LTRak detektagarriak zituzten ERVetan LTR bien analisi filogenetikoa burutu zen. Honetarako, MrBayes 3.1 programa erabilita (10^6 belaunaldi eta eredu ebolutiboa finkatu gabe), ERVen 5'-LTR eta 3'-LTR eskualdeen inferentzia bayestarrean oinarritutako zuhaitz filogenetikoa eraiki zen. Analisi hauetan taldez kanpoko moduan *Drosophilaren* ZAM elementuaren LTRak erabili ziren. Era berean LTRak detektagarriak zituzten ERVen parte zentralaren, hots, *gag+pol+env* geneen zuhaitz filogenetikoa eraiki zen.

Honetaz gain, detektatutako LTRak ERVen insertzio-denbora estimatzeko erabili ziren. Hau egiteko LTRen ezberdintasun kopurua erabili zen eta $2.3-5.0 \times 10^{-9}$ ordezen/leku/urte-ko tasa aplikatu zen insertzio-denbora kalkulatzeko (Johnson & Coffin, 1999).

IV.2.3.2. Hautespena zuhaitz filogenetiko mailan

Hautespena analizatzeko sekuentzia nukleotidikoek irakurtarua mantentzea beharrezkoa zen, hautatutako elementuen sekuentzietan insertzio/delezioen presentzia dela eta. Hautatutako elementuen BLAST bilaketaz identifikaturiko aminoazido-sekuentziak MAFFT (Katoh et al., 2005) programaren bidez (linsi aukera) lerrokatu ziren. Lerrokatzea honetan oinarrituz, elementu bakoitzaren nukleotidoak lerrokatu ziren, tripleteak zegokion aminoazidoaren posizioan kokatuz eta horrela irakurtarua mantenduz.

Hautatutako elementuen sekuentzia hauek PAML3.15 programarekin (Yang, 1997) analizatu ziren dn/ds proportzioa (ω) kalkulatzeko. ds-k ordezkapen sinonimo leku sinonimoko adierazten du eta dn-k ordezkapen ez-sinonimo leku ez-sinonimoko. PAML programak kalkulu hauek egiteko Goldman eta Yang-en eredu sinplifikatu bat erabiltzen du (Yang, 1997). Analisiak burutzeko PAML programari sekuentzia nukleotidikoekin eraikitako Maximum-Likelihood zuhaitzak eman zitzaizkion. Zuhaitz hauek Phyml programaren (Guindon & Gascuel, 2003) bidez eraiki zen zen eta ordezkapen-ereduak kalkulatzeko Modeltest (Posada & Crandall, 1998) programa erabili zen.

Programa honekin adar guztietan ω berdina onartzen duen eredia eta barneko eta kanpoko adarrek ω ezberdina dutela onartzen duen eredia kalkulatu ziren. Euren artean eredu probableena aukeratzeko egiantz-arrazoi testa (*likelihood ratio test*) burutu zen.

IV.2.4. ERVen sekuentzietan baimendutako aldaketen analisisa

ERVen sekuentzietan baimendutako aldaketak aztertzeke errekonbinazio intragenikoa eta sekuentzian zeharreko hautespena erabili ziren. Lehenak ERV ezberdinen arteko sekuentzien elkartrukeari buruzko informazioa eskaintzen digu eta bigarrenak sekuentzian mutazioak baimenduta ote daudenari buruzkoa.

IV.2.4.1. Errekonbinazio intragenikoa

Errekonbinazio intergenikoaren analisisian erabilitako kontserbatutako posizioez osotutako sekuentziak erabili ziren errekonbinazio intragenikoa aztertzeke. Analisi honetarako *Drosophila*-ren ZAM elementua ez zen kontutan hartu.

Errekonbinazio intragenikoa aztertzeke metodo ezberdinak erabiltzea gomendagarria da (Posada, 2002). Horretarako oinarri ezberdina duten bost metodo erabili ziren: bi filogenetiko, RDP (Martin & Rybicki, 2000) eta PLATO (Grassly & Holmes, 1997); eta ordezkapeneko hiru, Maxchi (Smith, 1992), Chimaera (Posada & Crandall, 2001) eta Geneconv (Sawyer, 1999). RDP, Maxchi eta Chimaera RDP3 programaren (Martin et al., 2005) bidez erabili ziren eta beste biak Geneconv (Sawyer, 1999) eta PLATO (Grassly & Holmes, 1997) erabilita.

Azkenengo programa honek, hots PLATOk, behar duen Maximum-Likelihood zuhaitza eraikitzeke Phyml programa (Guindon & Gascuel, 2003) erabili zen eta bai PLATOk zein Phyml-k

behar duten ordezkpan-ereduak kalkulatzeko Modeltest (Posada & Crandall, 1998) programa erabili zen.

Bai RDPn, zein Maxchi-n, Chimaera-n edota PLATOn konparazio anitzetarako Bonferroni zuzenketa erabili zen, emaitzen esangarritasunean zoriaren eragina ekiditeko. Geneconv-en, aldiz, Bonferroni zuzenketa erabili ahal izateko, lehengo 10000 permutazio egin ziren eta, eskaintzen zituen emaitza guztietatik, “Barneko zati globalak” (*Global inner fragments*) emaitzak kontutan hartu ziren, hots, gene konbertsioa aztertzeke erabilitako lerrokatzean zeuden sekuentzien artean estimatzen zuen emaitza eta esangarritasuna Bonferroni zuzenketaren bidez zuzentzen zituen emaitza mota.

RDP, Geneconv, Maxchi eta Chimaera metodoen errekonbinazio-gertaera eta erabilitako sekuentzia-kopuruaren arteko korrelazioa azterteko R lengoaia erabili zen (R Development Core Team, 2008).

IV.2.4.2. Haustespena sekuentzia-mailan

Gorago aipatu den moduan sekuentziak lerrokatu ostean, PAML programa erabiliz lau eredu ezberdin kalkulatu ziren sekuentzian zeharreko hautespena detektatzeko: M0 eredu (edo proportzio bakarrekoa), zeinetan sekuentzia osoan zehar eta adar guztietan ω bakarra dagoela onartzen den; M1 eredu (edo ia neutrala), zeinetan sekuentzian zehar $\omega < 1$ eta $\omega = 1$ duten lekuak daudela onartzen den; M2 eredu (edo positiboa), zeinetan sekuentzian zehar $\omega < 1$, $\omega = 1$ eta $\omega > 1$ duten lekuak daudela onartzen den; eta M3 eredu (edo diskreto), zeinetan sekuentzian zehar mugatu gabeko hiru ω ezberdin daudela onartzen den. Egiantz-arrazoi testen bidez M0 eta M3 artean eta M1 eta M2 ereduaren artean eredu probableena aukeratu zen ($p < 0,05$).

IV.3. Emaitzak

IV.3.1. ERVen detekzioa eta aukeraketa

Espezie bakoitzean hautatutako erretrobirus endogenoen (ERV) kopurua, hots, hiru gene erretrobiralak kontserbatuta zutenak eta, beraz, lan honetan burututako analisientzat egokiak direnak, espeziez espezie aldakorra izan zen. ERV hauek hainbat metafamilietan (Mf) sailkatu ziren espezie bakoitzeko ERVen *gag+pol+env* zuhaitz filogenetikoan oinarrituta (IV-3. taula).

Metafamiliek analisisetan sakontzea ahalbidetu zuten. Aurretik definitutako ERVen familiak ahalik eta modurik barneratzailenean taldekatzea izan zen metafamilia hauek eraikitzeke irizpidea eta, ondorioz, metafamilia bakoitzak kide-kopuru ezberdina izan zuen (IV-3. taula). Analizatutako ERV gehienak I klasean barneratzen dira eta, gutxi batzuk, II klasean (IV-3. taula). ERV hauen antzinatasunari dagokionez (IV-3. taula), hein zabala azaldu zuten.

IV.3.2. ERVen kopia-kopuruaren emendioa

ERVen kopia-kopuruaren emendioa aztertzeke errekonbinazio intergenikoa eta ERVen zuhaitz filogenetikoetan barneko eta kanpoko adarrek zuten hautespena erabili ziren, gene erretrobiralek jatorri berdina zuten eta funtzionalak mantentzen ote ziren jakiteko.

IV.3.2.1. Errekonbinazio intergenikoa

Errekonbinazio intergenikoaren analisirako oinarritzko hiru gene erretrobiralen zuhaitz filogenetikoak eraiki ziren (IV-1. - IV-21. irudi gehigarriak), bai batera zein banatuta. Zuhaitz hauen arteko desadostasunak, hots, sekuentzien erlazio filogenetiko ezberdinak, errekonbinazio intergenikoko gertaeren adierazle gisa erabili ziren. Ahal izan zen kasuetan, LTRen eta geneen zuhaitzak eraiki ziren ERVen LTRek eta geneek jatorri berdina ote zuten jakiteko.

Gizakia

Gizakitik hautatutako ERVen zuhaitz filogenetikoetan metafamiliek sendotasun estatistikorik azaldu ez bazuten ere, gizakiaren ERV batzuen geneek jatorri ezberdina azaldu zuten. (IV-1. irudia).

Mf4 metafamilia *chr7_13* ERVen *env* genea ez zen metafamilia honetako gainontzeko ERVen *env* geneekin multzokatu. Mf2-ko *chr13_45* eta *chr8_137* ERVen *gag* geneak eta Mf1-eko *chrY_36*

Mamalioen erretrobirus endogenoen (ERV) dinamika ebolutiboa

ERVen *gag* geneak batera multzokatu ziren, euren jatorrizko metafamiliatik aldenduz. Mf6-n, hainbat ERVen *pol* geneek jatorri ezberdina izan zuten, gainontzeko ERVen *pol* geneekin ez taldekatuz. Azkenik Mf3 eta Mf5-ko ERVen geneek ez zuten historia ebolutibo ezberdinik azaldu.

LTRen eta zati zentralaren historia ebolutiboa aztertzerakoan (IV-5. irudia), Mf4-ko hainbat ERVen bi LTRak eta Mf6-ko 19_4 ERVaren 3'LTR-ak ez ziren euren kideekin multzokatu. Bestalde, Mf1 eta Mf3-ko ERVen LTRak zuhaitzetan zehar sakabanatuak agertu ziren.

IV-3. taula: Aztertutako ERVen ezaugarriak espezie eta metafamilia bakoitzean.

Mf-z kanpo, metafamilia batean sailkatu ez diren ERVak. Erlazionaturiko familiariak ez bazegoen klasea adierazten da. Txertaketa-denbora orain dela milioi urtetan (ODMU) adierazi da. -, ezin izan da txertaketa-denbora estimatu.

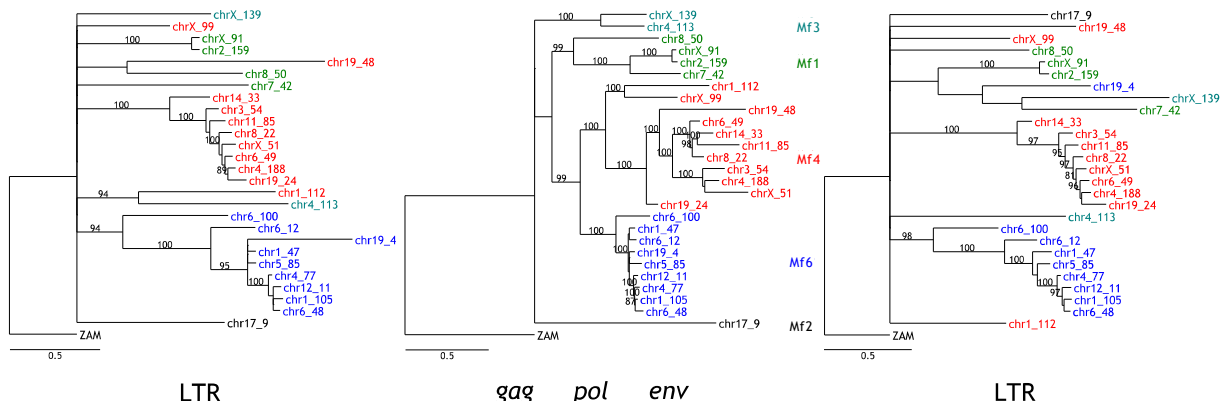
Espezia	Kopurua	Erlazionatutako familia (Klasea)	Txertaketa-denbora (ODMU)	
			Heina	Bataz-bestekoa
<i>Homo sapiens</i>	246			
Mf1	28	HERV-E/RRHERV-I (I)	8-33	19,37
Mf2	17	HERV-K (II)	15-33	24
Mf3	22	HERV-I/ADP (I)	19-41	30,75
Mf4	50	HERV-F/XA/H (I)	5-63	23,18
Mf5	3	(I)	-	-
Mf6	123	ERV9/HERV-W (I)	4-81	27,06
Mf- z kanpo	3	(I)	-	-
<i>Pan troglodytes</i>	157			
Mf1	6	CERV12 (I)	47-101	74
Mf2	5	(I)	-	-
Mf3	29	CERV6/7/8 (I)	1-53	17,1
Mf4	12	CERV30/32/36/37/39 (II)	14-55	31
Mf5	16	CERV20/22/23/24/25/26 (I)	-	-
Mf6	6	CERV29 (I)	-	-
Mf7	83	CERV17/18 (I)	1-35	13,5
<i>Mus musculus</i>	114			
Mf1	11	(I)	-	-
Mf2	5	(I)	-	-
Mf3	9	(I)	0-3	1,08
Mf4	8	(I)	0-14	3,56
Mf5	9	(I)	1-13	5,75
Mf6	24	Mmr3_MuRRS (I)	1-54	11,1
Mf7	16	(I)	0-43	15,78
Mf- z kanpo	32	(I)	-	-
<i>Rattus norvegicus</i>	59			
Mf1	7	(I)	-	-
Mf2	24	(I)	0.2-5	4,18
Mf3	6	Mmr3_MuRRS (I)	1-2	1
Mf4	6	(I)	0-13	6,6
Mf5	5	(I)	7-15	11
Mf- z kanpo	11	(I)	-	-
<i>Canis familiaris</i>	23			
Mf1	5	(I)	-	-
Mf2	9	(I)	11-25	18
Mf3	6	(I)	1-3	2
Mf-z kanpo	3	(I)	-	-
<i>Bos taurus</i>	93			
Mf1	8	BoERV19/23/24 (II)	23-50	36,5
Mf2	8	BoERV9/10 (I)	13-33	21
Mf3	8	BoERV7/12/14/15/16 (I)	13-57	31,25
Mf4	30	BoERV3 (I)	10-57	27,25
Mf5	39	BoERV1 (I)	0.2-40	7,59
<i>Equus caballus</i>	22			
Mf1	9	EqERV4 (I)	1-34	13,25
Mf2	7	EqERV1/2/3 (I)	14-30	22
Mf3	5	EqERV12/13/15 (II)	4-10	7
Mf-z kanpo	1	EqERV5 (I)	-	-



IV-1. irudia: Aukeratutako gizakiaren ERVen gag+pol+env geneen gene anizkoitzeko zuhaitz filogenetikoa.

Inferentzia Bayestarra, 10^6 belaunaldi, eredua finkatu gabe. Taldekapen hauek jarraitu ez duten ERVak adierazi dira eta alboan zein genetan kutxa barnean jarri dira.

Mamaloien erretrobirus endogenoen (ERV) dinamika ebolutiboa



LTR *gag* *pol* *env* LTR

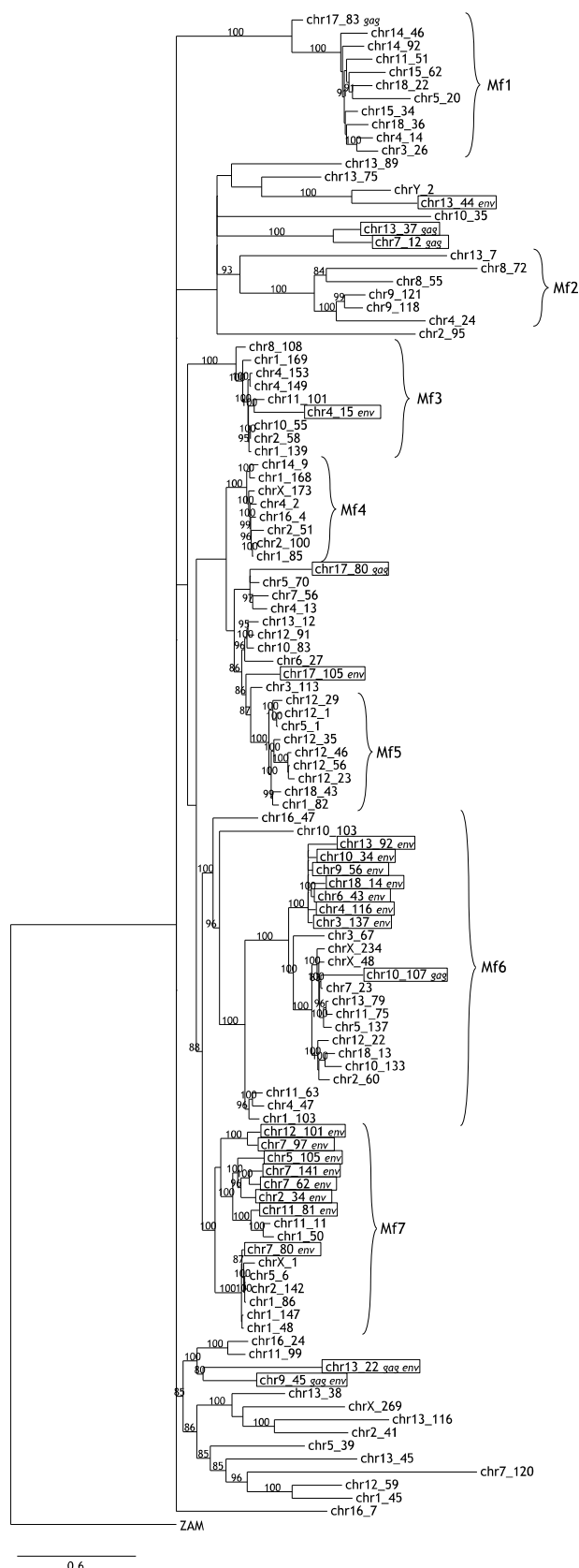
IV-2. irudia: LTRak detektatu zaien gizakien ERVen 5'LTR, parte zentrala eta 3'LTR zuhaitz filogenetikoak.

Ezkerrean 5'LTR, erdian parte zentrala (*gag+pol+env*) eta eskuman 3'LTR. Inferentzia Bayestarra, 10^6 belaunaldi, eredua finkatu gabea. Koloretan metafamilia bakoitzeko ERVak.



IV-3. irudia: Aukeratutako txinpantzearen ERVen *gag+pol+env* geneen gene anizkoitzeko zuhaitz filogenetikoak.

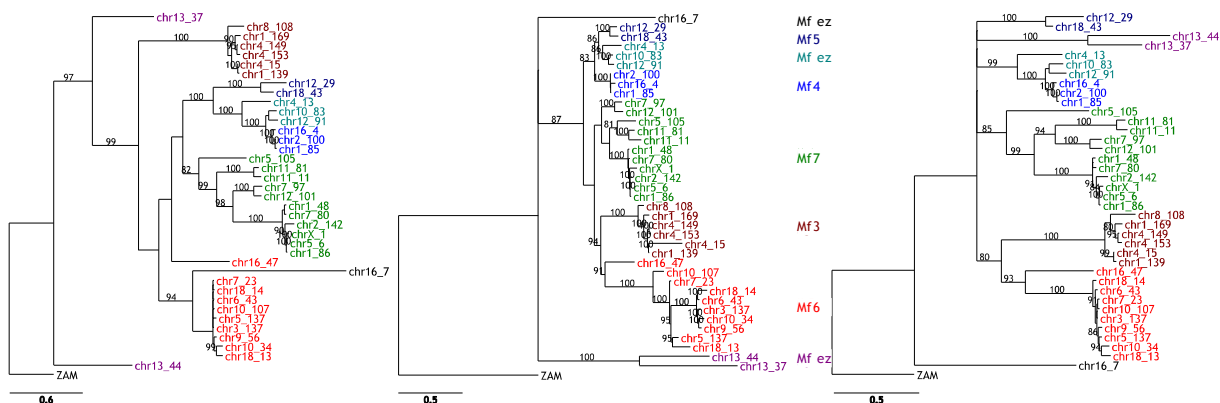
Inferentzia Bayestarra, 10^6 belaunaldi, eredua finkatu gabe. Taldekapen hauek jarraitu ez duten ERVak adierazi dira eta alboan zein genetari kutxa barnean jarri dira.



IV-4. irudia: Aukeratutako saguaren ERVen *gag+pol+env* geneen gene anizkoitzeko zuhaitz filogenetikoa.

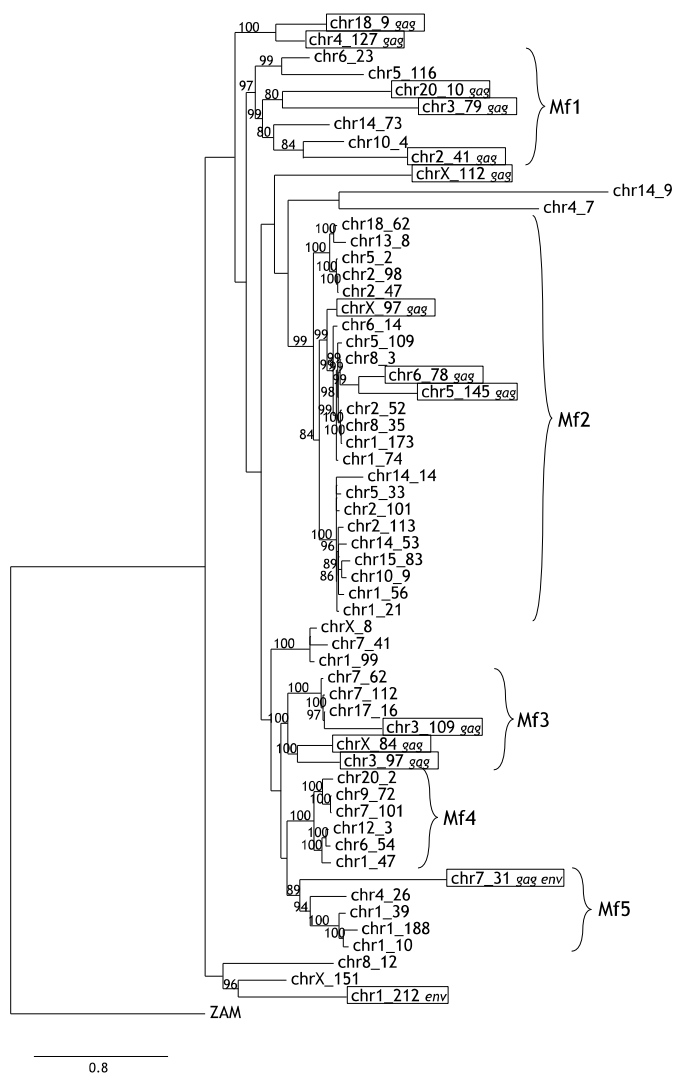
Inferentzia Bayestarra, 10^6 belaunaldi, erdua finkatu gabe. Taldekapen hauek jarraitu ez duten ERVak adierazi dira eta alboan zein genetari kutxa barnean jarri dira.

Mamalian retrovirus endogenous (ERV) dynamics evolution



IV-5. irudia: LTRak detektatu zaien saguen ERVen 5'LTR, parte zentrala eta 3'LTR zuhaitz filogenetikoak.

Ezkerrean 5'LTR, erdian parte zentrala (*gag+pol+env*) eta eskuman 3'LTR. Inferentzia Bayestarra, 10⁶ belaunaldi, eredu finkatu gabea. Koloretan metafamilia bakoitzeko ERVak.



IV-6. irudia: Aukeratutako arratoiaren ERVen *gag+pol+env* geneen gene anizkoitzeko zuhaitz filogenetikoa.

Inferentzia Bayestarra, 10^6 belaunaldi, erdua finkatu gabe. Taldekapen hauek jarraitu ez duten ERVak adierazi dira eta alboan zein genetari kutxa barnean jarri dira.

Txinpantzea

Txinpantzean erabilitako ERVetan, hainbat genek historia ebolutibo ezberdina izan zutela ikusi zen (IV-3. Irudia).

Mf1, Mf2 eta Mf6 metafamilietak hainbat ERVeren *env* geneak euren artean erlazio estuagoa azaldu zuten euren artean zegokien metafamiliakoekin baino. Mf3-n, hainbat *gag* gene eta chr17_12ko *env* genea ez ziren euren metafamiliakoekin multzokatu. Mf4-n, chr6_9 ERVaren *pol* genea metafamilia horren barruan beste kokapen batean agertu zen eta *env* genea Mf7-ko ERV batzuekin multzokatu zen. Mf5-eko chr14_3 ERVaren hiru geneak, bere kabuz aztertzerakoan, ez ziren modu berean taldekatu, gene bakoitzak erlazio ezberdina eduki baitzuen. Metafamilia honetan ere chr7_13 eta chr11_10 ERVen *gag* geneak beste ERV batzuekin elkartu ziren eta chr1_4ren *env* genea metafamiliako gainontzeko *env*etatik banatuta agertu zen. Azkenik, Mf7-n ez zen desadostasunik aurkitu.

Espezie honetan hainbat ERVeren LTRak identifikatu baziren ere, LTRen zuhaitz filogenetikoak ez zuen sendotasunik izan eta ez da aurkeztu.

Sagua

Geneak

Saguaren genomatik aukeratutako ERVen zuhaitz filogenetikoek hainbat genek jatorri ezberdina zutela adierazi zuten (IV-4. Irudia).

Metafamilietan ez sailkatutako hainbat ERVen geneek ere beste mota batzuetako erlazioak azaldu zituzten, bai *gag* zein *env* geneetan. Bestalde, Mf1 metafamiliako chr17_83-ren *gag* geneak metafamilia batean sailkatu gabeko chr9_45rekin erlazio estuagoa zuela zirudien, Mf3-ko chr4_15 ERVaren *env* genea ez zen bere metafamiliakoekin taldekatu eta Mf6-ko chr10_107 ERVaren *gag* geneak Mf3-koekin erlazio estuagoa zuela zirudien. Honetaz gain, Mf6 eta Mf7 metafamilietan, hainbat ERVen *env* geneak ez ziren gene horren zuhaitzean batera agertu, zuhaitzetik barreiatuta baizik. Azkenik, Mf2, Mf4 eta Mf5-en, aldiz, gene erretrobiralek ez zuten desadostasunik azaldu.

LTRak eta geneak

Sagan detektatu ziren LTRak eta ERV berberaren zati zentrala erkatuz, ez zen ezberdintasun handirik aurkitu (IV-5. irudia). Salbuespen bakarrak 5'LTR zuhaitzean eman ziren, non metafamilia batean sailkatu gabeko chr13_37 eta chr13_44 ERVen LTRak banatu ziren eta Mf6-ko chr16_47 LTRa ez zen gainontzeko Mf6-ko LTRekin multzokatu.

Arratoia

Geneak

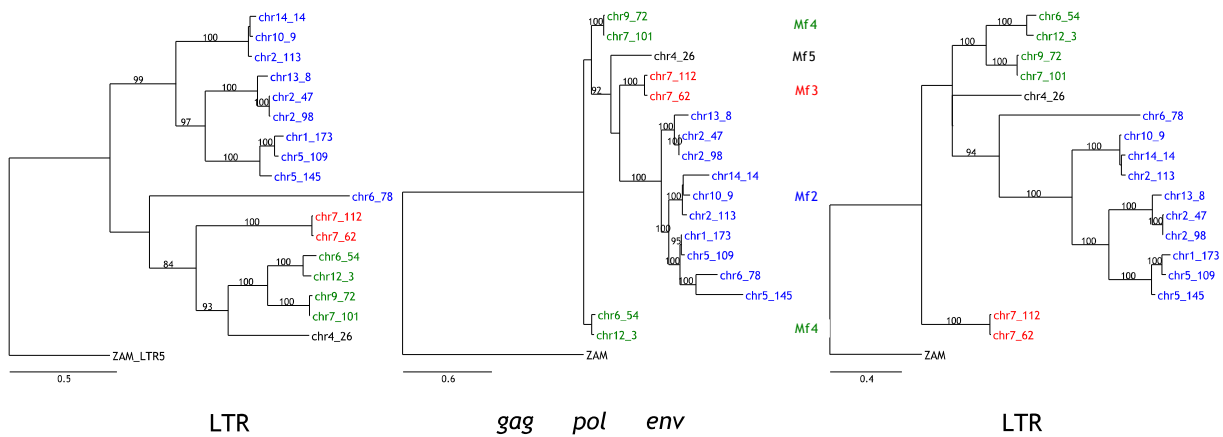
Arratoian analizatutako ERVen geneen artean, *gag* genearen filogeniak ezberdintasun gehien izan zituen izan zen (IV-6. irudia).

Mf1 metafamiliako chr3_79-ren *gag* genea Mf2-koen *gag* geneekin multzokatu zen eta metafamilia honetako chr2_41-en eta chr20_10-en eta Mf3-ko chr3_109-ren *gag* geneak batera agertu ziren. Mf2-ren barruan, chrX_97 ERVaren *gag* geneak erlazio ezberdina azaldu zuen eta chr5_145 ERVak Mf5-ko chr7_21 ERVaren *gag* genearekin erlazio estua izan zuen, gainontzekoetatik aldentuz. Bestalde, Mf2-ko chr6_78-ren *gag* geneak eta metafamilia batean sailkatu ez zen chrX_112-ren *gag* geneak erlazio estua azaldu zuten. Mf3-ko chrX_84-ren eta chr3_97-ren *gag* geneak gainontzeko metafamiliako kideen *gag* geneetatik urrundu ziren. *gag* geneaz gain, Mf5-eko chr7_31 ERVaren *env* genea ez zen metafamilia honetako gainontzeko *env* geneekin multzokatu.

LTRak eta geneak

LTRen eta zati zentralaren erlazioei dagokienez (IV-7. irudia), ez zen ezberdintasun handirik aurkitu. Soilik Mf2 metafamiliako barne-antolakuntza, geneekin erkatuz, LTRetan ezberdina izan zen. Bestalde, zati zentralaren zuhaitzean Mf4 metafamilia bi taldetan banatu zen, LTRetan gertatu ez zena.

Mamalioen erretrobirus endogenoen (ERV) dinamika ebolutiboa



IV-7. irudia: LTRak detektatu zaien arratoien ERVen 5'LTR, parte zentrala eta 3'LTR zuhaitz filogenetikoak.

Ezkerrean 5'LTR, erdian parte zentrala (*gag+pol+env*) eta eskuman 3'LTR. Inferentzia Bayestarra, 10^6 belaunaldi, eredua finkatu gabea. Koloretan metafamilia bakoitzeko ERVak.

Txakurra

ERVen gene ezberdinen zuhaitz filogenetikoek txakurraren hainbat ERVren geneak jatorri ezberdina zutela azaldu zuten (IV-8. Irudia).

Metafamilia batean sailkatu ez ziren ERVek, hots, chr8_3, chr10_1 eta chr20_4, *gag* genearen erlazio ezberdinak azaldu zituzten, lehenengoko bien *gag* geneak Mf2 metafamilian multzokatu ziren eta hirugarrenarena Mf3-ko chr1_10 eta chr35_4 ERVekin. Mf3-n, aipatutako elementuen *gag* genearen historia ezberdinaz gain, chr1_18 elementuaren *env* genea Mf3-ko ERVen *env* genearekin multzokatu beharrean, metafamilia batean sailkatu ez zen chr20_4 ERVen *env* genearekin batera agertu zen. Azkenik, Mf1 eta Mf2 metafamilien kasuan ez zen geneen historia ebolutibo ezberdinik detektatu.

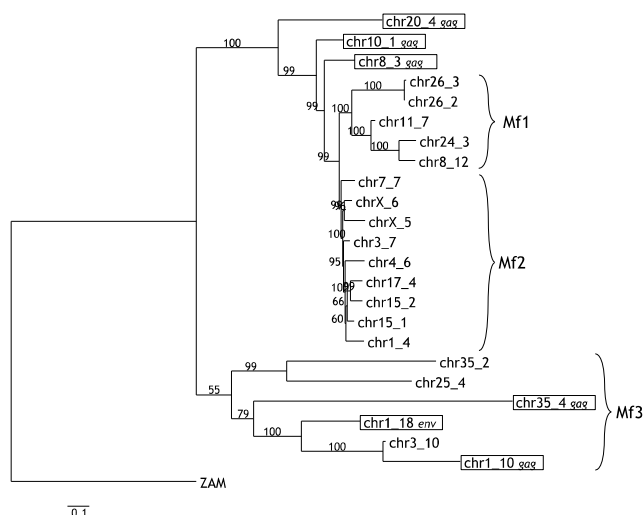
ERV hauetatik soilik 2 ERVen LTRak identifikatu ahal izan ziren eta ez zen zuhaitzik eraiki.

Behia

Geneak

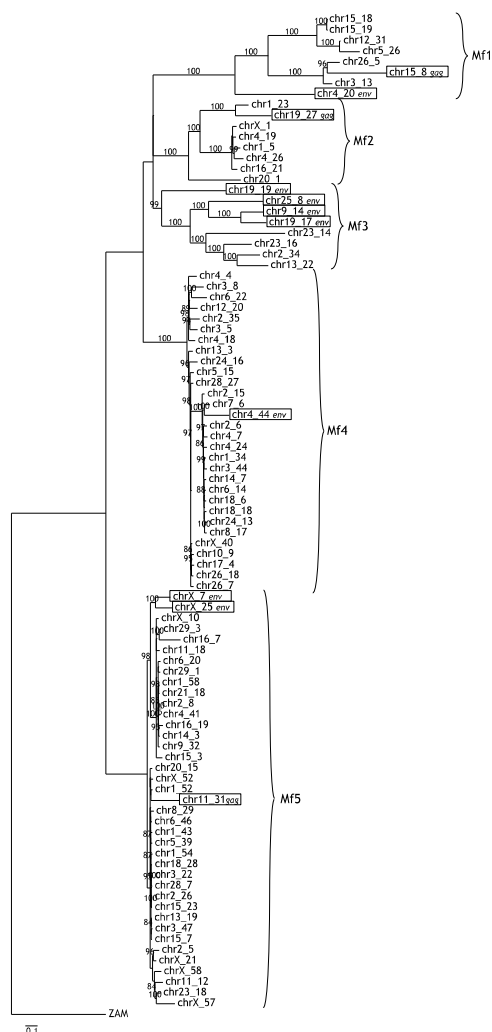
Analisi honetarako aukeratutako behiaren ERVen geneen zuhaitz filogenetikoek hainbat ERVren geneek historia filogenetiko ezberdina zutela azaldu zuten (IV-9. Irudia).

Mf1 metafamilien ERVen *env* genearen zuhaitz filogenetikoak erlazio esturik ez zuten bi multzo nagusi zeudela adierazi zuen eta chr15_8 ERVan *gag* genearen zuhaitz filogenetikotik gainontzeko ERVen *gag* geneetatik aldendu zela ondoriozta zitekeen. Mf2-n chr19_27 ERVaren *gag* genea ez zen metafamilia honetakoekin multzokatzen, baizik eta Mf5-ekoekin. Mf3-n chr19_19 eta chr25_8 ERVen *env* genea ez zen metafamilia honetako *env* geneekin ez gainontzeko metafamilien geneekin multzokatzen, are gehiago, chr19_19 ERVaren *env* genea gainontzeko ERVen *env* geneetatik aldendu zen. Mf4-n chr4_44 ERVaren *env* genea, gene honen zuhaitz filogenetikoa eraikitzerakoan, Mf3-ko *env* geneekin erlazioa zuela ikusi zen. Mf5-en, *gag* genearen zuhaitzean, chr11_31 ERVaren *gag* genea Mf3-tik gertuago zegoela metafamilia honetatik baino ikusi zen. chrX_7 eta chrX_25 ERVen *env* geneak ez ziren metafamilia bateko *env* geneekin multzokatzen gene honen zuhaitzean.



IV-8. irudia: Aukeratutako txakurraren ERVen *gag+pol+env* geneen gene-anizkoitzeko zuhaitz filogenetikoa.

Inferentzia Bayestarra, 10^6 belaunaldi, eredu finkatu gabe. Taldekapen hauek jarraitu ez duten ERVak adierazi dira eta alboan zein genetean kutxa barnean jarri dira.



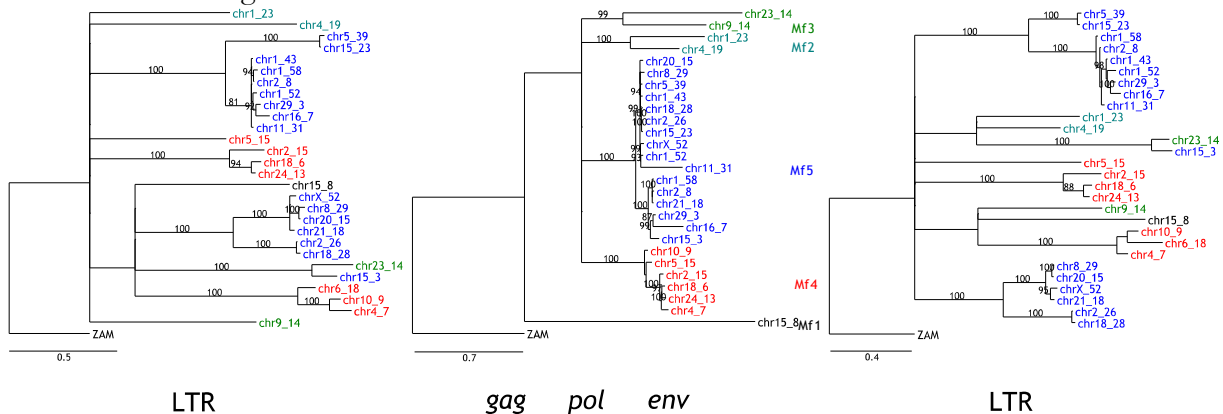
IV-9. irudia: Aukeratutako behiaren ERVen *gag+pol+env* geneen gene anizkoitzeko zuhaitz filogenetikoa.

Inferentzia Bayestarra, 10^6 belaunaldi, eredu finkatu gabe. Taldekapen hauek jarraitu ez duten ERVak adierazi dira eta alboan zein genetean kutxa barnean jarri dira.

Mamalioen erretrobirus endogenoen (ERV) dinamika ebolutiboa

LTRak eta geneak

LTRen eta zati zentralaren arteko erlazioari dagokionez (IV-10. irudia), Mf4 eta Mf5 metafamiliako LTR-ak bi taldetan banatu ziren. Gainera, Mf5-en kasuan, LTRen bi taldeak ez ziren bat etorri zati zentrolean ikusten ziren Mf5-eko bi azpitaldeekin. Baita ere, Mf5-eko chr15_3 ERVaren LTRak Mf3-ko 23_14 ERVaren LTRekin batera agertu ziren. Azkenik, Mf2-ko ERVen LTRak ez ziren esangarriki batera multzokatu.



IV-10. irudia: LTRak detektatu zaien behien ERVen 5'LTR, parte zentrala eta 3'LTR zuhaitz filogenetikoak.

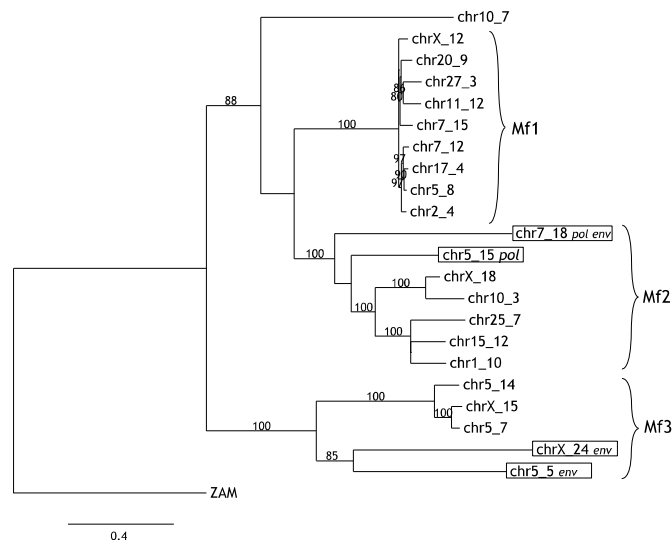
Ezkerrean 5'LTR, erdian parte zentrala (*gag+pol+env*) eta eskuman 3'LTR. Inferentzia Bayestarra, 10^6 belaunaldi, erdua finkatu gabea. Koloretan metafamilia bakoitzeko ERVak.

Zaldia

Zaldian aztertutako ERVen gene ezberdinen zuhaitz filogenetikoak oso argigarriak izan ez baziren ere, ERV batzuen geneek historia ezberdina azaldu zuten (IV-11. irudia).

Mf2 metafamilian multzokatu ziren chr5_15 eta chr7_18-ren *pol* geneak ez ziren gainontzeko Mf2-ko ERVen *pol* genearekin multzokatu gene horren zuhaitzean. chr7_18-ren kasuan, gauza bera gertatu zen *env* genearekin. Mf3-ko chr5_5 eta chrX_24 ERVei dagokionez, euren *env* geneak ez ziren metafamiliako gainontzeko ERVen *env* genearekin batera agertu *env* zuhaitzean.

Zaldian hautatutako ERVetan LTRak identifikatzea 4 kasutan posible izan zen. Hala ere, lortutako LTRen zuhaitzak ez ziren argigarriak izan eta ez dira erakusten.



IV-11. irudia: Aukeratutako zaldiaren ERVen *gag+pol+env* geneen gene anizkoitzeko zuhaitz filogenetikoa.

Inferentzia Bayestarra, 10^6 belaunaldi, erdua finkatu gabe. Taldekaper hauek jarraitu ez duten ERVak adierazi dira eta alboan zein genetan kutxa barnean jarri dira.

Laburpena

Espezie guztiak batera aztertuz, espeziez espezie jatorri komuna ez zuten ERVen gene-kopurua aldakorra izan zen (IV-4. Taula).

Oro har, *gag* eta *env* geneek *pol* baino errekonbinazio intergeniko gertaera gehiago azaldu zituzten, hots, osotzen zuten ERVaren gainontzeko geneen jatorri berbera ez zuten izan. *gag* genearen kasuan, oro har, 42 gertaera detektatu ziren, batez ere arratoian (13 gertaera, arratoian aztertutako ERVen %22,03a). *env* genean 49 gertaera detektatu ziren, batez ere saguan (20 gertaera, saguan aztertutako ERVen %14,54a) eta txinpantzean (16 gertaera, txinpantzean aztertutako ERVen %10,19a). *pol* genean, aldiz, 17 gertaera detektatu ziren soilik, gehienak gizakian (13 gertaera, gizakian aztertutako ERVen %5,28).

IV-4. taula: Errekonbinazio intergeniko gertaren laburpena, aztertutako filogenia ezberdinen geneetan.

N, aztertutako ERV kopurua; -, gertaerarik ez da detektatu.

Espeziea	N	Genea		
		<i>gag</i>	<i>pol</i>	<i>env</i>
Gizakia	246	4	13	1
Mf1	28	1	-	-
Mf2	17	2	-	-
Mf3	22	-	-	-
Mf4	50	-	-	1
Mf5	3	-	-	-
Mf6	123	1	10	-
Mf-z kanpo	3	-	3	-
Txinpantzea	157	10	2	16
Mf1	6	2	-	2
Mf2	5	-	-	3
Mf3	29	5	-	1
Mf4	12	-	1	1
Mf5	16	3	1	2
Mf6	6	-	-	6
Mf7	83	-	-	1
Sagua	114	7	-	20
Mf1	11	1	-	-
Mf2	5	-	-	-
Mf3	9	-	-	1
Mf4	8	-	-	-
Mf5	9	-	-	-
Mf6	24	1	-	7
Mf7	16	-	-	8

Espeziea	N	Genea		
		<i>gag</i>	<i>pol</i>	<i>env</i>
Mf-z kanpo	32	5	-	4
Arratoia	59	13	-	2
Mf1	7	3	-	-
Mf2	24	3	-	-
Mf3	6	3	-	-
Mf4	6	-	-	-
Mf5	5	1	-	1
Mf-z kanpo	11	3	-	1
Txakurra	23	5	-	1
Mf1	5	-	-	-
Mf2	9	-	-	-
Mf3	6	2	-	1
Mf-z kanpo	3	3	-	-
Behia	93	3	-	6
Mf1	8	1	-	1
Mf2	8	1	-	-
Mf3	8	-	-	4
Mf4	30	-	-	1
Mf5	39	1	-	-
Zaldia	22	-	2	3
Mf1	9	-	-	-
Mf2	7	-	2	1
Mf3	5	-	-	2
Mf-z kanpo	1	-	-	-

IV.3.2.2. Zuhaitz filogenetikoaren zeharreko hautespena

Hautespena detektatzeko dn/ds proportzioa (ω) erabili zen: $\omega > 1$ denean, hautespen positiboa dagoela esaten da; $\omega = 1$, hautespena neutrala dela; eta $\omega < 1$, hautespena purifikatzailea dela. Adar guztietan ω berdina dela onartzen duen eredu eta kanpoko eta barruko adarrek ω dutela onartzen duena kalkulatu ziren eta egiantz-arrazoi testa baten bitartez espezie eta metafamilia bakoitzean eredu probableena aukeratu zen (IV-5. taula) egiantz-arrazoi testa erabilita (2. taula gehigarrian emaitzak aztertu daitezke).

Gizakia

Gizakian hautatutako ERVen gene eta metafamilia gehienetan kanpoko eta barneko adarrek ω ezberdina izatea onartzen duen eredu zuhaitz osoan zehar berdina onartzen duen eredu baino probableagoa izan zen (IV-5. taula). Salbuespenak Mf2 metafamiliako ω *gag* eta Mf4-ko *env* genea. Kasu guztietan $\omega < 1$ izan zen, eta barruko adarrek kanpokoek baino ω txikiagoa azaldu zuten, Mf2-ko *env* genean izan ezik. Mf5-en, sekuentzia-kopurua dela eta, ez da analisi hau buru.

IV-5. taula: ERVen hautespenaren estimazioa aztertutako espezie, metafamilia eta geneetan. Adar guztietan ω berdina ($M0$ eredua) izatea eta barneko eta kanpoko adarrek ω ezberdina izatearen artean eredu probableena aztertzen duen analisiaren laburpena.

Testatutako bi ereduaren artean probableena azaltzen da: $\omega_i = \omega_t$ zutabea adar guztietan ω bakarra detektatu denean erabiltzen da; $\omega_i \neq \omega_t$ zutabea kanpoko eta barneko adarrek ω ezberdina dutenean. N: aztertutako sekuentzia-kopurua; ω , d_N/d_S balioa; ω_i , barneko adarren (inner branches) d_N/d_S balioa; ω_t , kanpoko adarren (terminal branches) d_N/d_S balioa.

Espeziea	N	Genea								
		gag			pol			env		
		$\omega_i = \omega_t$	$\omega_i \neq \omega_t$	ω_t	$\omega_i = \omega_t$	$\omega_i \neq \omega_t$	ω_t	$\omega_i = \omega_t$	$\omega_i \neq \omega_t$	ω_t
Gizakia	246	-	0,0824	0,3616	-	0,0789	0,3932	-	0,0852	0,2465
Mf1	28	-	0,0601	0,2703	-	0,0597	0,3108	-	0,0000	0,1673
Mf2	17	0,2355	-	-	-	0,0056	0,3577	-	0,3986	0,0883
Mf3	22	-	0,0410	0,2449	-	0,0335	0,3394	-	0,0439	0,1929
Mf4	50	-	0,1205	0,2754	-	0,0788	0,3314	0,2074	-	-
Mf5	3	-	-	-	-	-	-	-	-	-
Mf6	123	-	0,1198	0,3996	-	0,1032	0,4236	-	0,0798	0,2749
Txinpantzea	157	-	0,0765	0,3673	-	0,0874	0,2983	-	0,0313	0,2384
Mf1	6	0,3079	-	-	0,3592	-	-	0,0241	-	-
Mf2	5	0,1865	-	-	0,0635	-	-	0,1652	-	-
Mf3	29	-	0,0369	0,1906	0,1475	-	-	-	0,0240	0,1864
Mf4	12	0,1503	-	-	0,1008	-	-	-	0,3139	0,0230
Mf5	16	-	0,0276	0,3714	-	0,0641	0,2762	-	0,0175	0,1487
Mf6	6	-	319,1000	0,0727	-	999,0000	0,0482	0,0489	-	-
Mf7	83	-	0,1170	0,4250	-	0,1044	0,3994	-	0,3609	0,1676
Sagua	114	-	0,0791	0,2911	-	0,0602	0,3201	-	0,0658	0,1426
Mf1	11	-	0,1328	0,4117	0,4312	-	-	-	0,1837	0,6531
Mf2	5	-	0,0112	0,3374	-	42,7889	0,1589	-	181,6337	0,1049
Mf3	9	0,0563	-	-	-	1,6159	0,0255	-	0,0001	0,1112
Mf4	8	0,2272	-	-	0,3723	-	-	0,3213	-	-
Mf5	9	-	0,0445	0,3145	-	0,0660	0,2982	-	0,0394	0,2619
Mf6	24	-	0,1709	0,3919	-	1,2012	0,1682	-	0,0193	0,0559
Mf7	16	-	0,0461	0,1871	-	0,0375	0,1708	0,0771	-	-
Arratoia	59	-	0,0811	0,2133	-	0,0529	0,1303	-	0,1297	0,0691
Mf1	7	1,3891	-	-	-	0,0461	0,3900	-	0,0000	0,2368
Mf2	24	1,1010	-	-	-	0,0456	0,0940	-	0,1230	0,0717
Mf3	6	1,0830	-	-	-	0,0205	0,0698	0,0053	-	-
Mf4	6	1,2696	-	-	0,0705	-	-	0,0575	-	-
Mf5	5	1,0780	-	-	0,0197	-	-	0,0421	-	-
Txakurra	23	0,3062	-	-	0,3055	-	-	0,2855	-	-
Mf1	5	0,3118	-	-	0,3862	-	-	-	0,0815	999,0000
Mf2	9	-	0,0004	0,4715	0,3066	-	-	0,2325	-	-
Mf3	6	0,0307	-	-	0,0905	-	-	-	0,0001	0,0339
Behia	93	-	0,0740	0,2915	-	0,0738	0,4612	-	0,0557	0,1767
Mf1	8	0,1531	-	-	-	0,0857	0,2652	-	0,0300	0,2898
Mf2	8	-	0,0022	0,3312	-	0,0400	0,7911	0,1348	-	-
Mf3	8	0,1007	-	-	-	0,0149	0,3305	-	16,9273	0,0230
Mf4	30	0,2688	-	-	-	0,1985	0,6402	0,2019	-	-
Mf5	39	-	0,1417	0,3060	0,2148	-	-	0,1075	-	-
Zaldia	22	-	0,0244	0,3157	-	0,2894	0,0100	-	0,4634	0,1144
Mf1	9	-	0,0130	0,1983	0,1505	-	-	0,1736	-	-
Mf2	7	-	0,0012	0,5135	0,1301	-	-	-	0,0004	0,3171
Mf3	5	-	999,0000	0,1544	0,2461	-	-	-	0,0076	0,3829

Txinpantzea

Txinpantzean aztertutako ERVek zuhaitz osoan zehar ω berdina zuten ala kanpoko eta barneko adarrek ω ezberdina azaltzen zuten aztertzerakoan (IV-5. taula), genez gene eta metafamiliaz metafamilia emaitza ezberdinak lortu ziren. Salbuespenak salbuespen, $\omega < 1$ izan zen eta, kanpoko adarrek eta barnekoek ω ezberdina azaldu zutenean, barneko adarren ω txikiagoa izan zen. Orokortasun hauetatik kanpo Mf4 eta Mf7 metafamilietako *env* genea eta Mf6-ko *gag* eta *pol* gelditu ziren. Kasu hauetan barneko adarren ω handiagoa izan zen eta Mf6-ko *gag* eta *pol* geneetan kanpoko adarrek $\omega > 1$ azaldu zuten.

Sagua

Hautatutako saguaren ERVen geneek eta metafamiliek, oro har, kanpoko eta barneko adarren ω

ezberdina zutela azaldu zuten (IV-5. taula). Hala ere, adar guztiek ω bakarria zutela Mf1 mefamiliako *pol* geneek, Mf3-ko *gag*ek, Mf4-ko hiru geneek, eta Mf7-ko *env*ek azaldu zuten. Honetaz gain, kanpoko eta barneko ω ezberdinak izan zirenean, kasu gehienetan barneko adarrek ω txikiagoa azaldu zuten, Mf2, Mf3 eta Mf6-ko *pol*en eta Mf2-ko *env*en izan ezik. Gainera, kasu hauetan, barneko adarren ω kanpokoena baino handiagoa izateaz gain, beren balioa bat baino handiagoa izan zen.

Arratoia

Kanpoko eta barneko adarrek ω berdina ala ezberdina zuten aztertzerakoan aztertutako arratoiaren ERVen geneetan eta metafamilietan emaitzak aldakorak izan ziren (IV-5. taula). *gag* genearen kasuan, sekuentzia guztiak aztertzeko orduan, barneko adarrek kanpokoek baino ω txikiagoa azaldu zuten eta biak $\omega < 1$ izan ziren. Metafamilietan, aldiz, *gag* geneak ω bakarria azaldu zuen eta hau bat baino handiagoa izan zen. *pol* genearen kasuan, Mf4 eta Mf5 metafamilietan izan ezik, barneko eta kanpoko adarrek ω ezberdina azaldu zuten eta, gainera, barneko adarrek ω txikiagoa azaldu zuten. Azkenik, *env* genean, aztertutako sekuentzia guztiek eta Mf1-ek barneko adarrek kanpokoek baino ω txikiagoa zutela azaldu zuten, Mf2-k, aldiz, kontrakoa eta Mf3, Mf4 eta Mf5-k zuhaitzean zehar ω bakarria.

Txakurra

Kanpoko eta barneko adarretan ω bakarria ala ezberdina al zen aztertzerakoan (IV-5. taula), hautatutako txakurraren ERVen gene eta metafamilia gehienetan ω bakarria onartzen duen eredia probableagoa izan zen. Salbuespenak Mf1 eta Mf3 metafamiliako *env* geneak eta Mf2-ko *gag* izan ziren. Kasu guztietan $\omega < 1$ izan zen, Mf1-ko *env* geneko zuhaitzaren kanpo adarrean izan ezik. Bestalde, kanpoko eta barneko adarretan ω ezberdina izan zenean, barneko adarrek ω txikiagoa azaldu zuten.

Behia

Behiaren ERVen sekuentzia guztiak edota metafamiliak aztertzerakoan, barneko eta kanpoko adarren ω -ei dagokionez (IV-5. taula), berdina zirenean ω bat baino txikiagoa izan zen eta ezberdinak zirenean, barneko adarrek kanpokoek baino ω txikiagoa izan zuten. Salbuespen bakarria Mf3-ko *env* genea izan zen, zeinetan barneko ω 1 baino handiagoa izan zen eta kanpoko adarren ω baino handiagoa.

Zaldia

Barneko eta kanpoko adarren ω -ei dagokionez (5. taula), hautatutako zaldiaren ERVen sekuentzia guztiak aztertzerakoan, ω guztiak bat baino txikiagoak izan ziren. *gag* genearen kasuan barneko adarren ω kanpokoena baino txikiagoa izan zen eta *pol* eta *env* geneetan, aldiz, alderantziz. Metafamiliari dagokionez, oro har, barneko adarren ω kanpokoena baino txikiagoa izan zen. Hala ere salbuespenak egon ziren: Mf1 metafamiliako *env* genea, zeinetan kanpoko eta barneko adarrek ω berdina izatea probableagoa izan zen, eta Mf3 metafamiliako *gag*, zeinetan barnekoko adarrek kanpokoek baino ω handiagoa izan zuten.

Laburpena

Espezie guztiak batera hartuz, hainbat kasutan zuhaitz filogenetikoaren adar guztietan zehar ω berdina izatea ezberdina izatea baino probableagoa izan zen (IV-5. taula). Barneko eta kanpoko adarrek ω ezberdina izatea probableagoa izan zen kasuetan, barneko adarrek kanpokoek baino ω txikiagoa azaldu zuten, salbuespenak salbuespen. Hortaz, analisi honetan oinarrituta eta Belshaw et al. (2008)-ren irizpidea jarraituta, espezie gehienek ERVetan, kopia-kopuruaren emendioa berrinfekzioaren bidez gertatu dela ondorioztatu daiteke. Orokortasun hauetatik at arratoiaren metafamilietan sailkatzen ziren ERVen *gag* genea egon zen. Gene honetan kasu guztietan $\omega > 1$ izan zen eta, hortaz, arratoiaren ERVetan berrinfekzioa baztertu daiteke.

IV.3.3. ERVen sekuentzietan baimendutako aldaketak

ERVen sekuentzietan baimenduta dauden aldaketetan sakontzeko errekonbinazio intragenikoa eta sekuentzian zehar gertatzen den hautespena aztertu ziren, sekuentzien arteko informazio trukerik

zegoen eta mutazioak pairatzea baimenduta ote zegoen jakiteko.

IV.3.3.1. Errekonbinazio intragenikoa

Errekonbinazio intragenikoa detektatzeko bost metodo ezberdin erabili ziren (RDP, Maxchi, Chimaera, Geneconv eta PLATO) eta bakoitzetik informazio ezberdinak lortu ziren. Informazio guztia IV-6. taulan laburbiltzen da. Metodo bakoitzaren emaitza zehatzak 3. taula - 9. taula gehigarrietan aztertu daitezke.

Gizakia

Erabilitako giza-ERV guztiekin analisiak egiterakoan, erabilitako metodo ezberdinek hiru geneen sekuentzietan zehar errekonbinazio-gertaerak detektatu zituzten, baina ez adostasun-maila berarekin. *gag* genearen kasuan metodo gehienek sekuentzia osoan zehar detektatu zuten errekonbinazioa; *pol* genean, aldiz, hasierako eta bukaerako posizioak eskualde errekonbinanteak izan zitezkeela aditzera eman zuten; azkenik, *env* genean, 200. posizioaren inguruan izan ezik, errekonbinazioa sekuentzia osoan zehar detektatu zuten.

Metafamiliak aztertzerakoan, *gag* genean, Mg4 metafamilian 300. posiziotik bukaera arte errekonbinazio-gertaerak detektatu zituzten metodo ezberdinek. Mf3-n, aldiz, 400. posiziotik aurrera adostasuna murriztu zen. Mf2-ren kasuan lehengoko 100 posizioetan metodo gehienek errekonbinazioa detektatu zuten; Mf1-en heina 200. posiziorarte zabalduz eta azkenengo posizioetan eskualde errekonbinante bat detektatuz eta Mf6 metafamilian, 200. posiziotik bukaera arte. Azkenik, Mf5 metafamilian seinalea ahulagoa izan zen, 50.etik 200.era bitarteko posizioak nabarmenduz.

pol geneari dagokionez, Mf1, Mf3 eta Mf6 metafamilien kasuan azkenengo 150 posizioek eratzen duten eskualdean errekonbinazioa detektatu zen eta Mf2-n 400. posizioaren inguruan. Mf4-n, batez ere lehengoko 100 posizioetan eta 300.etik 600.era bitarteko posizioetan errekonbinazioa detektatu zuten. Azkenik Mf5-n, seinaleak ahulagoak izan ziren.

env genean, Mf1 metafamilian metodoen arteko adostasunak gutxi izan ziren. Mf2-n, ordea, 250. posiziotik aurrera seinalea sendoagoa izan zen eta Mf3-n 300. posiziotik aurrera. Mf4 eta Mf6-n ia sekuentzia osoan zehar errekonbinazioa gertatu zela bat etorri ziren metodo ezberdinak, Mg6-n 100.etik 200.era bitarteko posizioetan izan ezik. Azkenik, Mf5-n errekonbinazio-gertaera gutxi detektatu ziren, sendoena 200.etik 300.era bitarteko posizioetan dagoen eskualdean izan zen.

Txinpantzea

Aukeratutako txinpantzearen ERV guztiak aztertzerakoan, *gag* genean 200. posizioaren inguruan eta 450. posiziotik aurrera metodo gehienek errekonbinazioa detektatu zuten, *pol* genearen kasuan 500. posiziotik aurrera eta *env*-n 200. posiziotik aurrera.

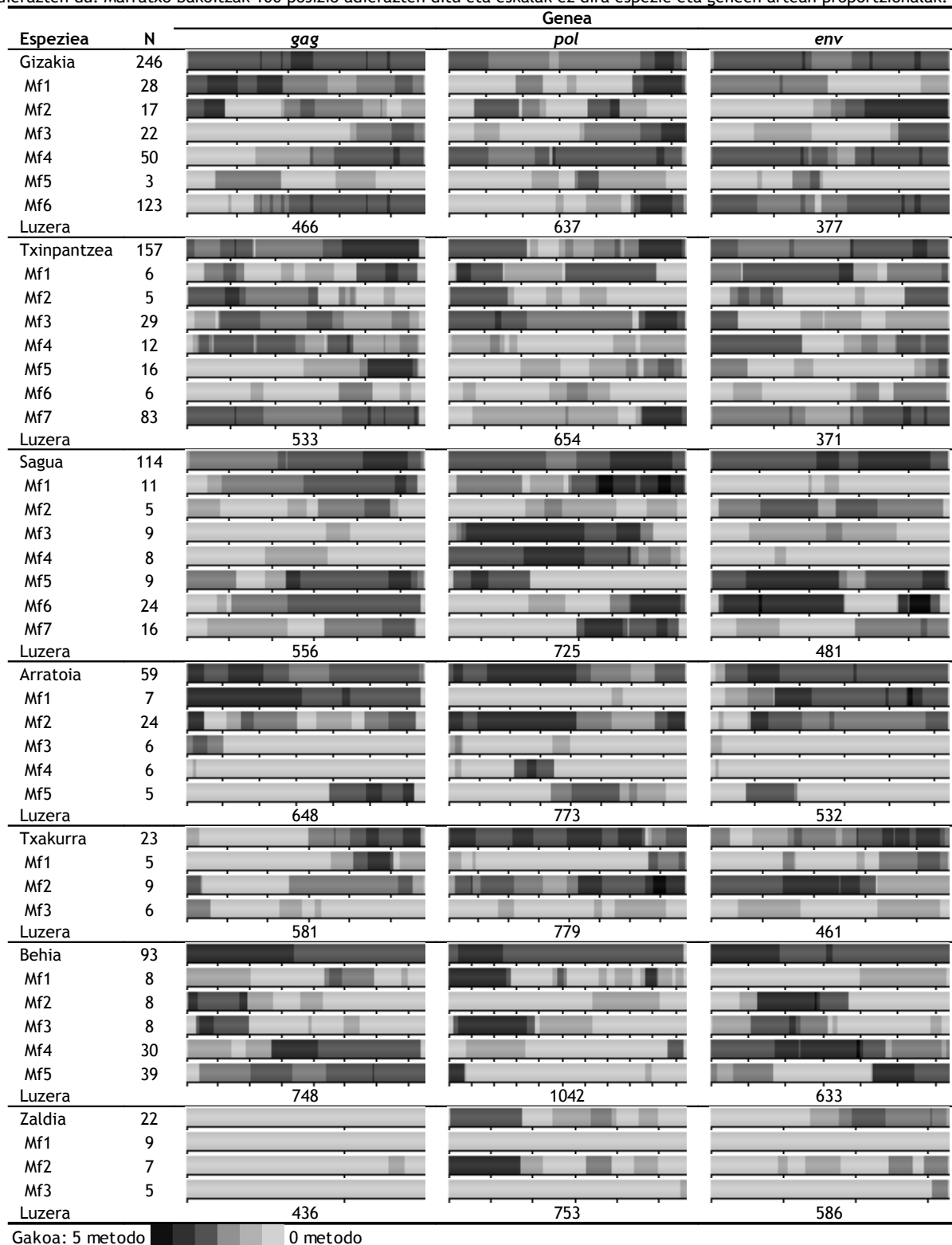
Metafamiliak aztertuz, *gag* genean Mf1 metafamilian 400.etik 500.era bitarteko posizioetan, Mf2-n lehenengoko 100 posizioetan eta 300. posizioaren inguruan, Mf3-n 100. eta 300. posizioen inguruan, Mf4-n 50.etik 250.era bitarteko posizioetan eta 300.etik 400.era bitarteko posizioetan, Mf5-n 400.etik 500.era bitarteko posizioetan eta Mf7-n 1.etik 150.era eta 350.etik 500.era bitarteko eskualdean errekonbinazio gertaeren detekzioan bat etorri ziren metodo gehienak. Mf6-n, ordea, seinaleak ahulak izan ziren.

pol genearen kasuan, adoste nabarmenenak Mf1 metafamiliako 50.etik 150.era eta 300.etik 550.era bitarteko posizioetan, Mf2-ko lehengoko 150. posizioetan, Mf3 metafamiliako 1.etik 200.era bitarteko eskualdean eta 600. posizioaren inguruko eskualdean eta Mf7-ko 600. posizioaren inguruan eman ziren. Gainontzeko metafamilietan adostasunak ez ziren horren argiak izan, metodoen batek errekonbinazio-gertaerak detektatzen bazituen ere.

env geneari dagokionez, metodoen arteko adostasunak aldakorrak izan ziren. Metodo gehienek bat egin zuten eskualdeak honakoak izan ziren: Mf1 metafamilian 50.etik 220.era bitarteko posizioetan, Mf2-n 50.etik 100.era bitartekoetan, Mf3-n hasierako eskualdean, Mf4-n hasierako eskualdean, 300. posizioaren inguruan eta azkenengo 100 posizioetan, Mf5-en hasierako posizioetan eta azkenengo 100 posizioetan, eta Mf7-n azkenengo 150 posizioetan. Mf6-n, aldiz, adosteak ez ziren horren argiak izan.

IV-6. taula: ERVen errekonbinazio intragenikoaren detekzioa aztertutako espezie, metafamilia eta geneetan.

Posizio jakinaren kolore beltzaren intentsitateak posizio horretan errekonbinazioa detektatu duten metodo kopurua adierazten du. Marratxo bakoitzak 100 posizio adierazten ditu eta eskalak ez dira espezie eta geneen artean proportzionalak.



Sagua

Hautatutako saguaren ERV guztietan errekonbinazioa aztertzerakoan, *gag* genean 400.etik 500.era bitarteko posizioetan errekonbinazio gertaeren presentziarekin bat etorri ziren metodo gehienak; *pol* dagokionez 500.etik 700.era bitartekotan, eta *env* 200.etik 250.era eta 300.etik 450.era bitartekotan.

Metafamiliaz metafamilia aztertuz, *gag* genean seinalerik sendoenak Mf1 metafamilian 300.etik 550.era bitarteko posizioetan, Mf2-n 400.etik 500.era bitartekoetan, Mf5-n eta Mf6-n 150.etik 550.era bitartekoetan eta Mf7-n 500. posizioaren inguruan eman ziren. Mf3 eta Mf4-n, ordea, seinaleak ahulak izan ziren.

pol genearen kasuan, Mf1 metafamilian 450.etik 700.era bitarteko posizioetan, Mf3-n 50.etik 600.era bitartekoetan, Mf4-n 1.etik 550.era bitartekoetan, Mf5-en 1.tik 250.era bitartekoetan, Mf6-n 650.etik 700.era bitartekoetan eta Mf7-n 400.etik 700.era bitartekoetan errekonbinazio-gertaerak eman zirela metodo ezberdinak bat etorri ziren. Mf2-n, aldiz, ez zen adostasun argirik eman.

Azkenik, *env* genean, metodo ezberdinek Mf5 (1.etik 250.era bitarteko posizioetan eta bukaerako eskualdean) eta Mf6 (50.etik 250.era bitartekoetan eta 400. posizioaren inguruan) metafamilietan errekonbinazioaren detekzioan bat egin zuten. Metodo ezberdinen adostea Mf2, Mf3 eta Mf7-n txikiagoa izan zen eta Mf1 eta Mf4-n ez zen ia errekonbinaziorik detektatu.

Arratoia

Arratoiaren ERVetan gertatutako errekonbinazio-gertaerak aztertuz, hiru geneen sekuentzian zehar metodo ezberdinek modu ezberdinean detektatu zuten errekonbinazioa. *gag* genean, batez ere errekonbinanteak hasierako eskualdeak ziruditen eta 100.etik 200.era bitarteko posizioak, *pol* kasuan hasierako posizioak eta 150.etik 400.era bitartekoak, eta *env* 100. posizioaren inguruan.

Metafamilia aztertzerakoan, *gag* geneari dagokionez, Mf1 metafamilian sekuentzia ia osoan zehar errekonbinazioa detektatu zen, batez ere hasierako 300 posizioetan eta 400. posiziotik gertu. Halaber, Mf2-ko hasierako posizioetan eta 600. posizioaren inguruan metodo gehienek errekonbinazioa detektatu zuten. Mf5-en 400.etik 600.era bitarteko eskualdean errekonbinazioa gertatu zela metodo gehienek bat egin zuten. Mf3-n aldiz, seinalea ahulagoa izan zen eta Mf4-n ez zen ia errekonbinaziorik detektatu.

pol genearen kasuan, Mf1 metafamilian eta Mf3-n errekonbinazio-gertaera gutxi detektatu ziren. Gainontzeko metafamilietan, ordea, errekonbinazio-gertaerak detektatu zituzten metodo gehienek: eMf2-n lehengoko 400 posizioetan eta azkenengo 100 posizioetan, Mf4-n 200.etik 350.era bitarteko eskualdean eta Mf5-en 400.etik 550.era bitartekoan.

Azkenik, *env* genean, Mf1 metafamilian 150. posiziotik bukaera arte errekonbinazio-gertaerak metodo gehienek detektatu zituzten, baita Mf2-n 100.etik 200.era bitarteko posizioen inguruan, 300. posiziotik gertu eta bukaerako posizioetan eta Mf5-en 100.etik 200.era bitarteko eskualdean. Mf3 eta Mf4-n, aldiz, ia ez zen errekonbinaziorik detektatu.

Txakurra

Txakurran hautatutako ERV guztiak kontutan hartuz, *gag* genean metodo gehienak bat datoz 400 posiziotik aurrera errekonbinazio-gertaerak detektatzerakoan. *pol* genean, aldiz, ia sekuentzia osoan errekonbinazio-gertaerak metodo ezberdinek detektatu zituzten. Azkenik, *env* genean sekuentzian zehar programa ezberdinek modu aldakorreen eskualde errekonbinanteak detektatu bazituzten ere seinalea azkenengo 150 posizioetan argiagoa izan zen.

Metafamilia bakoitza aztertuz, Mf1 metafamilian seinalerik sendoena *gag* genearen azkenengo posizioetan eman zen eta gainontzeko geneetan metodo ezberdinak posizio gutxitan bat etorri ziren. Mf2-n, aldiz, gainontzeko metafamilietan baino eskualde errekonbinante gehiago detektatu zituzten metodo ezberdinek. Seinale hauek *pol* genean, 400.etik 500.era bitarteko posizioetan eta 700. posizioaren inguruan sendoagoak izan ziren eta *env* genean 150.etik 300.era bitarteko eskualdean. Azkenik, Mf3 metafamilian, eskualde errekonbinanteen seinalea oso ahula izan zen hiru geneetan, metodo ezberdinak ez baitziren bat etorri gertaerak detektatzerakoan.

Behia

Analizatutako behiaren ERV guztiak aztertzerakoan, metodo gehienak hiru gene erretrobiraletan errekonbinazio-gertaerak detektatu zituzten. *gag* genearen kasuan, batez ere lehengo 300 posizioetan metodo ezberdinek errekonbinazioa eman zela ondorioztatu zuten; *pol* genean 50.etik 250.era bitarteko posizioetan eta 300. posizioaren inguruan; eta *env* genean lehenengoko 200 posizioetan eta 300. posizioaren inguruan.

Metafamiliak aztertuz, errekonbinazioaren detekzioa aldakorra izan zen. *gag* geneari dagokionez, Mf1 metafamilian adoste gutxi egon ziren metodoen artean. Aipagarriena 450.etik 500.era bitarteko eskualdea izan zen, zeinetan metodo ezberdinek errekonbinazioa detektatu zuten. Mf2 eta Mf3-n ere adostasunek hein murriztua azaldu zuten: 1.etik 200.era bitartekoetan eta 50.etik 100.era bitartekoetan hurrenez hurren. Mf4-n, ordea, metodo gehienak 300.etik 700.era bitartekoetan errekonbinazio-gertaerak egon zirela detektatu zuten, batez ere 300.etik 400.era bitartekoetan. Azkenik, Mf5-n 300.etik 400.era bitarteko posizioetan eta 550tik bukaera arte, eskualde errekonbinanteak izatearekin bat etorri ziren metodo ezberdinak.

pol genearen errekonbinazio analisian, Mf1 metafamilian 1.etik 250.era bitarteko posizioetan eta 900. posizioaren inguruan errekonbinazioa gertatu izanarekin metodo gehienak bat etorri ziren, Mf3-n 50.etik 400.era bitartekoetan, Mf4-n 1000. posizioaren inguruan eta Mf5-en hasierako ia 100 posizioetan. Mf2-n, ordea, metodo bakarrak detektatu zuen errekonbinazioa.

Azkenik, *env* genean, Mf1 metafamilian metodo bakarrak detektatu zuen errekonbinazioa eta Mf3-n seinalea ahula izan zen, 100.etik 250.era bitarteko posizioak gehiago nabarmenduz. Mf2-n, aldiz, metodo gehienek 150.etik 300.era bitartekoetan errekonbinazioa detektatu zuten eta Mf4-n 200.etik 400.era bitartekoetan, metodo ezberdinek sekuentzia zehar errekonbinazioa gertatu zela modu aldakorrean detektatu bazuten ere. Azkenik, Mf5-en metodo gehienek 400. posiziotik bukaera arte errekonbinazio-gertaerak detektatu zituzten.

Zaldia

Zaldian analizatutako ERV guztiak aztertuz, *gag* genean ez zen errekonbinaziorik detektatu. Beste bi geneetan, aldiz, seinaleak sendoagoak izan ziren: *pol*-en 1.etik 150.era bitarteko posizioetan eta *env*-en 250.etik 250.era bitartekoetan.

Mf2 metafamiliaren *pol* genean 1.etik 150.era bitarteko posizioetan metodo gehienek errekonbinazioa detektatu zuten eta neurri txikiago batean *pol*-en 300.etik 400.era bitartekoetan eta *env*-en bukaerako eskualdean. Mf1 eta Mf3 metafamilietan, aldiz, ez zen errekonbinaziorik detektatu, Mf3-ko *env* genearen azkenengo posizioetan izan ezik.

Laburpena

Errekonbinazio intragenikoa espezie eta metafamilia bakoitzean modu ezberdinean gertatu zela ikusi zen (IV-6. taula). Espezie eta metafamilia gehienetan, errekonbinazio seinale sendoak detektatu ziren eskualde batean edo sekuentzia osoan. Batez ere gizakiaren, txinpantzearen, saguaren, arratoiaren eta behiaren aztertutako ERV guztiekin egindako analisisetan errekonbinazioa detektatu zen. Salbuespena aztertutako zaldiaren ERVak izan ziren, zeintzuetan gertaera gutxi detektatu ziren, bai sekuentzia guztiak erabilia zein metafamilietan.

Geneak aztertuz, *gag* genean, hainbat metafamilia (adibidez, saguaren Mf3 eta Mf4 eta arratoiaren Mf4) eta zaldia izan ezik, errekonbinazio-gertaera nahiko detektatu ziren. *pol* genean, aldiz, espezie guztietan errekonbinazioa detektatu bazen ere gainontzeko bi geneetan baino gertaera gutxiago detektatu ziren, salbuespenak salbuespen. Azkenik, *env* genean errekonbinazio-gertaera nahiko detektatu ziren eskualde luzeagotan, hainbat metafamilietan izan ezik, hala nola, arratoiaren Mf3-a eta Mf4-a edo Zaldiaren Mf1-a.

Bestalde, errekonbinazio intragenikoko gertaerak detektatzeko erabilitako sekuentzia-kopuruak zuen eragina aztertu zen. Horretarako, RDP, Geneconv, Maxchi eta Chimaera metodoetan erabilitako sekuentzia-kopurua eta detektatutako errekonbinazio-gertaeren arteko korrelazioa burutu zen. Analisi honetatik PLATO kanpo gelditu zen gertaerak beharrean eskualde errekonbinanteak

detektatzen baititu. RDP metodoan izan ezik ($r=0,2609$; $p=0,0028$) gainontzeko metodoetan ez zen korrelazio esangarririk detektatu (Geneconv $p=0,04244$; Maxchi $p=0,0989$; Chimaera $p=0,9967$).

IV.3.3.2. Hautespena sekuentzian zehar

Hautespena detektatzeko dn/ds proportzioa (ω) erabili zen: $\omega > 1$ denean, hautespen positiboa dagoela esaten da; $\omega = 1$, hautespena neutrala dela; eta $\omega < 1$, hautespena purifikatzailea dela. Sekuentzia zeharreko ω kalkulatzeko 4 eredu aztertu ziren: (i) sekuentzia zehar ω bakarra dagoela (M0); (ii) sekuentzia zehar bi ω daudela, bata $\omega < 1$ eta bestea $\omega = 1$ (hautespen ia neutrala, M1); (iii) sekuentzia zehar hiru ω daudela, $\omega < 1$, $\omega = 1$ eta $\omega > 1$ (hautespen positiboa, M2); eta (iv) sekuentzia zehar hiru ω ezberdin daudela, aurretik balioa tartea finkatu gabe (M3). Egiantz-arrazoi test baten bitartez (emaitzak 2. taula gehigarrian aztertu daitezke), espezie eta metafamilia bakoitzean ω bakarra eta 3 ω ereduaren artean probableena zein zen kalkulatu zen (IV-7. taula) eta hautespen ia neutrala eta hautespen positibo artean probableena zena kalkulatzeko (IV-8. taula).

Gizakia

Hautatutako gizakien ERVen gene eta metafamilia gehienetan sekuentzia zehar hiru ω egotea bakarra egotea baino probableagoa izan zen (IV-7. taula). Salbuespen bakarrak Mf5 metafamiliako *gag* genea eta Mf2 eta Mf5-eko *env* izan ziren. Hala ere, kasu guztietan sekuentziaren kodoi gehienak hautespen purifikatzailepean egon ziren. Bestalde, hautespen positiboa ere aztertutako sekuentzia guztien *gag* eta *env* geneetan, Mg1 metafamiliako hiru geneetan, Mf2 eta Mf3ko *gagen*, Mf1-eko *enven*, Mf5-ko *enven* eta Mf6-ko *gag* eta *polen* detektatu ziren, kodoien %1,954tik %26,314ra bitartean.

Hautespen ia-neutrala eta positiboaren artean probableena zein izan zen kalkulatzeko orduan, genez gene eta metafamiliaz metafamilia emaitza aldakorra izan zen (IV-8. taula). Hala ere kodoi gehienak hautespen purifikatzailepean egon ziren. *gag* genean, hautespen positiboan zeuden kodoiak aztertutako sekuentzia guztietan (kodoien %0,328), Mf2 (%0,828), Mf4 (%0,832) eta Mf6 (%6,264) metafamilietan detektatu ziren. *pol* genearen kasuan, hautespen positiboan aztertutako sekuentzia guztien kodoien %0,399 Mf4-ko kodoien %3,600a, Mf5-ko %26,314a eta Mf6-ko %1,657a egon ziren. *env* genean, aldiz, ez zen hautespen positiboan zeuden kodoirik detektatu aztertutako sekuentzia guztietan izan ezik (%1,051). Azkenik, hautespen neutralpean kodoien %5,671etik %29,435era bitartean egon ziren.

Txinpantzea

Aztertutako txinpantzearen ERVen gene eta metafamilia gehienetan sekuentzia zehar 3 ω ezberdin egotea bakarra baino probableagoa izan zen (IV-7. taula), Mf1 eta Mf2 metafamilietako *env* genean eta Mf6-ko *polen* izan ezik. Kasu guztietan, kodoi gehienek hautespen purifikatzailea azaldu zuten. Bestalde, aztertutako sekuentzia guztien, Mf1, Mf4 eta Mf7 metafamilietako *gag* genean, Mf1 eta Mf7-ko *polen* eta Mf4 eta Mf7-ko *enven* hautespen positiboan zeuden kodoiak detektatu ziren, kodoien %2,813tik %17,976ra bitartean.

Aztertutako txinpantzearen ERVen gene eta metafamilia gehienetan hautespen neutrala positiboa baino probableagoa izan zen (IV-8. taula). Salbuespenak aztertutako sekuentzia guztien *gag* (kodoien %0,232a) eta *pol* (%0,157), Mf1-eko *pol* (%6,468), Mf4-ko *gag* (%0,629) eta *env* (%0,979) eta Mg7-ko *gag* (%2,917) eta *pol* (%0,430) izan ziren. Honetaz gain, kodoi gehienak hautespen purifikatzailepean egon ziren eta hautespen neutrala kodoien %5,097tik %34,892ra bitartean detektatu zen.

Sagua

Hautatutako saguaren ERVen gene eta metafamilia gehienetan sekuentzia zehar 3 ω egotea bakarra baino probableagoa izan zen (IV-7. taula), Mf2 metafamiliako *gag* genean eta Mf3-ko *enven* izan ezik. Sekuentziaren kodoi gehienetan hautespena purifikatzailea izan bazen ere, hautespen positiboan zeuden kodoiak ere detektatu ziren (Mf1 eta Mf4-ko hiru geneetan, Mf2-ko *polen* eta *enven*, Mf3-ko *polen* eta Mf5 eta Mf6-ko *gagen* eta *polen*). Hautespen positiboan zeuden kodoien proportzioa %2,806tik %33,502ra bitartekoa izan zen.

IV-7. taula: ERVen hautespenaren estimazioa aztertutako espezie, metafamilia eta geneetan. ω bakarra egotearen (M0 eredua) eta 3 ω ezberdin egotearen (M3 eredua) artean eredu probableena estimatzen duen analisiaren laburpena.

ω_b adierazten ez den kasuetan eredu probableena M0 da, ω_b adierazten denean M3 eredua probableena da. N: aztertutako sekuentzia-kopurua; ω_o : sekuentzia osoaren ω (d_N/d_S) balioa; ω_b : sekuentzian zehar dauden ω balio ezberdinen banakatzea; p: ω_b bakoitzaren proportzioa sekuentzia osoarekiko.

Especiea	N	Genea								
		gag			pol			env		
		ω_o	ω_b	p (%)	ω_o	ω_b	p (%)	ω_o	ω_b	p (%)
Gizakia	246	0,2974	0,13829 0,35517 1,45959	43,171 53,580 3,249	0,3121	0,13171 0,29195 0,58417	28,748 48,576 22,677	0,3092	0,13578 0,34305 10,29723	70,080 28,801 1,119
Mf1	28	0,1784	0,01976 0,11465 0,34876	10,516 58,004 31,480	0,2540	0,07638 0,23817 0,59955	34,815 45,223 19,962	0,1875	0,03342 0,15897 1,37516	34,443 59,656 5,902
Mf2	17	0,3154	0,03038 0,28996 4,56183	22,331 75,715 1,954	0,2650	0,09856 0,44445 0,90920	64,871 25,465 9,664	0,1334	- - -	- - -
Mf3	22	0,2334	0,05791 0,21395 1,34795	43,013 49,356 7,632	0,2616	0,04820 0,16526 0,56329	26,703 53,208 20,089	0,1393	0,05039 0,12504 0,40134	35,222 50,110 14,668
Mf4	50	0,2969	0,11165 0,46725 2,33542	61,024 36,479 2,497	0,3769	0,12332 0,47584 3,11650	62,460 32,949 4,592	0,2491	0,09399 0,26563 1,06395	37,444 56,573 5,983
Mf5	3	0,0659	- -	- -	15,2497	0,65842 32,25556	73,686 26,314	0,1679	- -	- -
Mf6	123	0,4140	0,14062 0,44994 1,76061	45,145 46,946 7,909	0,4426	0,14133 0,48310 1,62864	33,251 60,366 6,383	0,2571	0,09385 0,26326 1,25802	49,323 42,897 7,780
Txinpantzea	157	0,2784	0,13257 0,34532 1,53118	47,149 50,038 2,813	0,2555	0,10008 0,22494 0,44929	25,853 46,127 28,020	0,1861	0,05334 0,13917 0,31886	13,100 54,512 32,388
Mf1	6	0,6682	0,12933 1,81314 1,81316	67,996 26,690 8,314	1,0420	0,00000 0,33179 4,92432	34,758 47,267 17,976	0,0241	- - -	- - -
Mf2	5	0,2220	0,06344 0,35649 -	45,882 54,118 -	0,2292	0,00000 0,09214 0,72588	14,596 61,691 23,740	0,1652	- - -	- - -
Mf3	29	0,1513	0,01309 0,10684 0,31361	14,492 57,416 28,092	0,1701	0,02948 0,13781 0,42695	28,188 50,085 21,727	0,1585	0,04241 0,15552 0,73524	47,710 42,475 9,815
Mf4	12	0,4568	0,02246 0,22949 13,77975	40,855 56,843 2,301	0,1227	0,01268 0,08399 0,22817	18,431 58,752 31,135	0,3288	0,03637 0,10678 4,01653	36,185 57,485 6,331
Mf5	16	0,2264	0,03141 0,11850 0,49522	12,963 55,408 31,630	0,2037	0,02712 0,14638 0,47046	10,354 68,159 21,487	0,1283	0,01680 0,10259 0,63468	26,409 64,496 9,095
Mf6	6	0,1589	0,04127 0,60445 -	79,102 20,898 -	0,1297	- - -	- - -	0,1873	0,00000 0,05774 0,65743	5,903 71,920 22,177
Mf7	83	0,4367	0,16719 0,66463 3,25641	66,270 29,806 3,924	0,4254	0,11978 0,47541 1,39875	36,362 55,053 8,585	0,2665	0,09028 0,29286 1,44874	50,291 43,179 6,530
Sagua	114	0,1901	0,07219 0,17582 0,52238	26,180 61,880 11,940	0,2172	0,09574 0,22555 0,48909	41,540 41,170 17,290	0,1272	0,05004 0,11211 0,28899	32,093 48,127 19,781
Mf1	11	0,4680	0,16185 1,00000 -	63,478 36,522 -	0,5505	0,15906 0,70588 1,83136	48,504 41,733 9,763	0,9226	0,00000 0,48032 2,98467	26,149 51,176 22,675
Mf2	5	0,1052	- -	- -	1,5537	0,18498 5,27437	73,107 26,893	0,7423	0,06375 2,08905	66,498 33,502
Mf3	9	0,0690	0,00029 0,03165 0,83559	81,193 10,996 7,810	0,2180	0,00140 1,37708 -	84,256 15,744 -	0,0008	- - -	- - -
Mf4	8	0,2900	0,01025 0,23078 3,71712	20,371 76,640 2,988	0,5144	0,00000 0,42487 4,13076	32,288 61,595 6,117	0,5559	0,00000 0,47288 8,56613	45,918 50,373 3,709
Mf5	9	0,2485	0,03481 0,21713 1,01449	42,724 43,577 13,699	0,2751	0,05894 0,46142 1,34552	57,915 36,791 5,294	0,1493	0,12921 0,95206	97,559 2,441

Mamalioen erretrobirus endogenoen (ERV) dinamika ebolutiboa

Espezia	N	Genea								
		gag			pol			env		
		ω_o	ω_b	p (%)	ω_o	ω_b	p (%)	ω_o	ω_b	p (%)
Mf6	24	0,3825	0,00100 0,25541 2,04271	10,165 81,274 8,560	0,3677	0,00740 0,17465 1,27884	15,933 64,166 19,901	0,0544	0,01074 0,04879 0,26561	56,923 30,516 12,561
Mf7	16	0,1251	0,01966 0,20212 0,86736	67,338 25,770 6,892	0,1028	0,04558 0,37762	82,756 17,244	0,0857	0,01099 0,05618 0,25216	19,769 60,624 19,607
Arratoia	59	0,1474	0,04540 0,15424 0,39061	36,540 49,530 13,930	0,0873	0,03427 0,09998 0,30742	50,223 39,980 9,797	0,1114	0,03280 0,08105 0,25590	30,365 43,890 25,745
Mf1	7	7,4785	1,06663 2,77075 71,34174	53,029 38,788 8,183	1,6285	0,06077 0,35072 16,84829	31,696 60,002 8,302	0,0204	0,00688 0,04674	66,195 33,805
Mf2	24	1,4148	0,46353 1,49874 8,03274	35,369 60,312 4,319	0,1427	0,02610 0,35429 8,24353	84,233 14,945 0,822	0,1233	0,02987 0,30957 0,30958	66,599 16,868 16,533
Mf3	6	10,6024	0,18616 1,25939 638,06553	7,806 90,714 1,480	0,0534	0,00000 0,11784	54,709 45,291	0,0053	- - -	- - -
Mf4	6	3,9674	1,26959 4,55882 279,82502	97,833 1,222 0,945	0,0880	0,02320 0,54864	87,668 12,332	0,0649	0,00000 0,05391 0,69982	27,321 68,695 3,984
Mf5	5	4,1518	0,47152 1,53384 96,55570	35,797 61,047 3,155	0,0197	- - -	- - -	0,0421	- - -	- - -
Txakurra	23	0,3881	0,14791 0,54552 1,59285	55,074 39,053 5,874	0,3578	0,11096 0,32588 1,73960	19,890 74,824 5,283	7,5018	0,05215 0,29357 178,99487	15,901 80,046 4,054
Mf1	5	3,8031	0,00000 0,45757 79,58673	37,827 57,726 4,447	0,4620	0,17481 1,01703	65,899 34,101	0,5430	- - -	- - -
Mf2	9	0,3787	0,00000 0,43799 1,37097	25,065 69,527 5,409	0,4491	0,00000 0,09305 1,27349	22,380 45,694 31,926	1,5251	0,02962 0,58345 67,29865	65,422 32,623 1,955
Mf3	6	0,2886	0,00482 0,07509 1,17021	32,078 46,365 21,557	0,2827	0,00000 0,09250 1,18016	8,154 73,667 18,179	8,8392	0,00000 0,31900 156,73757	9,385 85,148 5,466
Behia	93	0,2917	0,1137 0,28238 15,23161	39,311 60,177 0,512	0,3283	0,19073 0,50868	56,734 43,266	0,4167	0,06875 0,21449 27,90434	53,789 45,191 1,013
Mf1	8	0,2022	0,02343 0,29954	35,249 64,751	0,2319	0,00000 0,10925 0,36683	6,889 42,580 50,531	0,8926	0,04421 0,15201 42,42326	36,874 61,280 1,846
Mf2	8	1,7398	0,06301 0,51154 24,07668	72,009 21,408 6,583	0,4387	0,0429 0,38426 1,40072	15,692 73,705 10,603	0,2768	0,00000 0,06072 1,94844	4,987 83,408 11,605
Mf3	8	1,0151	0,02449 0,13270 23,65872	38,886 57,184 3,930	0,1987	0,06508 0,30215 0,74863	58,378 33,787 7,834	0,0462	0,00814 0,02866 0,07948	15,667 43,437 40,896
Mf4	30	0,3719	0,15343 0,92866 20,06788	84,852 14,620 0,528	0,7004	0,43497 1,71707 8,21475	85,290 13,527 1,183	0,5479	0,00000 0,06728 1,14766	3,395 51,903 44,702
Mf5	39	0,3350	0,17074 0,76681 5,42812	83,531 15,051 1,418	0,2716	0,09123 0,63345 4,37246	78,617 19,659 1,724	0,1075	0,00000 0,06072 1,94844	4,987 83,408 11,605
Zaldia	22	0,1805	0,00534 0,09948 0,30441	3,582 55,261 41,158	0,1860	0,05726 0,26446	37,862 62,138	0,2384	0,04275 0,16569 1,03003	9,362 80,889 9,749
Mf1	9	0,2752	0,04992 1,11838 -	78,918 21,082 -	1,5229	0,00000 0,20941 151,63813	27,076 72,020 0,905	0,2403	0,05350 0,73051 1,26291	75,046 21,605 3,349
Mf2	7	0,2461	0,02698 0,33164 -	28,091 71,909 -	0,1734	0,00000 0,07086 0,34470	8,719 51,565 39,716	0,2222	0,01389 0,13345 1,13445	17,123 71,967 10,910
Mf3	5	0,3721	0,00000 0,08613 0,84007	29,208 29,526 41,266	2,7353	0,05579 0,24118 24,41195	36,884 52,514 10,602	0,5914	0,00003 0,07150 2,74298	16,193 63,911 19,896

IV-8. taula: ERVen hautespenaren estimazioa aztertutako espezie, metafamilia eta geneetan. Eredu ia neutralaren (M1 eredua) eta hautespen positiboaren (M2 eredua) artean eredu probableena estimatzen duen analisiaren laburpena.

Bi ω_b balio adierazten badira eredu probableena M1 da, hiru ω_b balio adierazten badira M2 eredu probableena da. N: aztertutako sekuentzia-kopurua; ω_o : sekuentzia osoaren ω (d_N/d_S) balioa; ω_b : sekuentzian zehar dauden ω balio ezberdinen banakatzeara; p: ω_b bakoitzaren proportzioa sekuentzia osoarekiko.

Espezia	N	Genea								
		gag			pol			env		
		ω_o	ω_b	p (%)	ω_o	ω_b	p (%)	ω_o	ω_b	p (%)
Gizakia	246	0,5026	0,23972	85,562	0,4360	0,26522	86,487	0,3760	0,17890	90,950
			1,00000	14,110		1,00000	13,114		1,00000	7,999
			47,67080	0,328		18,91614	0,399		12,67668	1,051
Mf1	28	0,1876	0,13873	94,329	0,3336	0,18162	81,432	0,1848	0,09447	90,022
			1,00000	5,671		1,00000	18,568		1,00000	9,978
Mf2	17	0,6425	0,17734	69,736	0,3386	0,13711	76,648	0,2357	0,09230	84,196
			1,00000	29,435		1,00000	23,352		1,00000	15,804
			27,10018	0,828						
Mf3	22	0,2254	0,11979	88,005	0,2616	0,14868	86,738	0,1874	0,10792	91,088
			1,00000	11,995		1,00000	13,262		1,00000	8,912
Mf4	50	1,0638	0,16409	76,228	0,4450	0,17797	79,937	0,2577	0,18345	90,901
			1,00000	22,940		1,00000	16,462		1,00000	9,099
			85,25194	0,832		3,83761	3,600			
Mf5	3	0,1072	0,04748	93,736	8,9730	0,65842	73,686	0,2971	0,13639	81,385
			1,00000	6,264		32,25556	26,314		1,00000	18,615
Mf6	123	0,8351	0,22810	78,063	0,4880	0,26456	76,760	0,2726	0,15456	88,182
			1,00000	21,218		1,00000	21,583		1,00000	10,782
			61,82101	0,720		4,17075	1,657		2,75296	1,036
Txinpantzea	157	0,5636	0,22170	90,319	0,5793	0,22176	92,206	0,1973	0,15418	94,903
			1,00000	9,449		1,00000	7,637		1,00000	5,097
			115,93634	0,232		190,12040	0,157			
Mf1	6	0,4450	0,09861	61,572	2,1351	0,04503	58,640	0,2962	0,03983	73,300
			1,00000	38,428		1,00000	34,892		1,00000	26,700
						27,20737	6,468			
Mf2	5	0,3486	0,13758	75,534	0,3921	0,06898	65,299	0,3584	0,10947	72,047
			1,00000	24,466		1,00000	34,701		1,00000	27,953
Mf3	29	0,2010	0,13249	92,102	0,2338	0,12874	87,941	0,2277	0,08253	84,182
			1,00000	7,898		1,00000	12,059		1,00000	15,818
Mf4	12	1,6629	0,08732	73,694	0,1995	0,09815	88,759	0,3791	0,06816	90,232
			1,00000	25,677		1,00000	11,241		1,00000	8,790
			213,28862	0,629					23,46484	0,979
Mf5	16	0,3010	0,13996	81,270	0,2377	0,15788	90,517	0,1578	0,07912	91,455
			1,00000	18,730		1,00000	9,483		1,00000	8,545
Mf6	6	0,2202	0,05184	82,240	0,2166	0,09416	86,483	0,2647	0,05148	77,522
			1,00000	17,760		1,00000	13,517		1,00000	22,478
Mf7	83	0,4996	0,20714	75,266	0,5347	0,22320	69,258	0,2541	0,15495	88,265
			1,00000	21,817		1,00000	30,311		1,00000	11,735
			4,30352	2,917		17,89601	0,430			
Sagua	114	0,3141	0,16723	91,976	0,5305	0,18883	86,959	0,4410	0,10789	94,312
			1,00000	7,454		1,00000	12,840		1,00000	4,713
			15,02905	0,571		118,25396	0,201		29,94379	0,976
Mf1	11	0,4680	0,08771	42,121	0,5244	0,18682	58,485	1,0583	0,08219	43,681
			0,58479	55,076		1,00000	41,515		1,00000	39,381
			6,47606	2,803					3,71103	16,939
Mf2	5	0,0881	0,00756	91,887	0,1691	0,00132	83,204	0,4286	0,03717	59,346
			1,00000	8,113		1,00000	16,796		1,00000	40,654
Mf3	9	0,0881	0,00756	91,887	0,1691	0,00132	83,204	0,0008	0,0008	100,0000
			1,00000	8,113		1,00000	16,796			
Mf4	8	0,2530	0,10677	83,633	0,7388	0,12164	69,847	0,6446	0,02914	66,585
			1,00000	16,367		1,00000	28,363		1,00000	30,730
						20,67929	1,790		11,84212	2,685
Mf5	9	0,2662	0,10961	82,414	0,3347	0,09946	83,879	0,1501	0,12952	97,632
			1,00000	17,586		1,00000	26,121		1,00000	2,368
Mf6	24	1,0107	0,21647	88,026	0,4154	0,11924	75,080	0,1153	0,03604	91,774
			1,00000	7,528		1,00000	22,546		1,00000	8,226
			16,75497	4,446		4,22746	2,374			
Mf7	16	0,1558	0,05447	89,279	0,1316	0,07074	93,449	0,1210	0,06855	94,363
			1,00000	10,721		1,00000	6,551		1,00000	5,637
Arratoia	59	0,1903	0,12348	92,371	0,1308	0,07389	93,853	0,1625	0,09946	92,997
			1,00000	7,629		1,00000	6,147		1,00000	7,003
Mf1	7	1,9608	1,00000	67,472	0,3669	0,16167	75,518	0,1432	0,01402	86,896
			3,95372	32,528		1,00000	24,482		1,00000	13,104

Mamalioren erretrobirus endogenoen (ERV) dinamika ebolutiboa

Especiea	N	Genea								
		<i>gag</i>			<i>pol</i>			<i>env</i>		
		ω_s	ω_b	p (%)	ω_s	ω_b	p (%)	ω_s	ω_b	p (%)
Mf2	24	1,1688	0,27686 1,00000 5,14506	16,823 76,170 7,006	0,1584	0,04079 1,00000 8,78791	94,066 5,154 0,780	0,1341	0,08521 1,00000	94,653 5,347
Mf3	6	10,5242	1,00000 604,72935	98,422 1,578	0,0717	0,04142 1,00000	96,844 3,156	0,0341	0,00463 1,00000	97,043 2,957
Mf4	6	1,0000	1,0000	100,0000	0,0971	0,03536 1,00000	93,600 6,400	0,0686	0,04015 1,00000	97,032 2,968
Mf5	5	1,3384	0,31943 1,00000 5,62629	13,636 77,042 9,321	0,0354	0,00783 1,00000	97,220 2,780	0,0420	0,0421	100,0000
Txakurra	23	0,4393	0,19494 1,00000	69,645 30,355	0,3846	0,25350 1,00000 5,30191	87,163 12,018 0,820	7,6080	0,19840 1,00000 183,62180	85,476 10,531 3,994
Mf1	5	0,4042	0,04910 1,00000	62,654 37,346	0,4598	0,17196 1,00000	65,233 34,767	0,6535	0,11932 1,00000	39,342 60,658
Mf2	9	0,4293	0,18647 1,00000	70,149 29,851	0,3934	0,04492 1,00000	63,509 36,491	1,9336	0,05882 1,00000 84,17482	74,981 23,048 1,971
Mf3	6	0,3678	0,03639 1,00000	65,602 34,398	0,2941	0,07490 1,00000	76,308 23,692	10,3921	0,16294 1,00000 204,21753	51,951 43,213 4,836
Behia	93	0,3143	0,19367 1,00000 15,79896	93,408 6,136 0,456	0,4195	0,28046 1,00000 12,81025	86,812 12,815 0,374	0,4075	0,11019 1,00000 28,68945	91,608 7,588 0,804
Mf1	8	0,5538	0,07400 1,00000	48,190 51,810	0,2944	0,18094 1,00000	86,150 13,850	0,1482	0,09645 1,00000	94,275 5,725
Mf2	8	2,6808	0,07269 1,00000 45,64826	78,216 16,395 5,389	0,4487	0,25523 1,00000	74,028 25,972	0,1906	0,05280 1,00000	85,452 14,548
Mf3	8	0,2113	0,07768 1,00000	85,514 14,486	0,2907	0,12360 1,00000	80,938 19,062	0,2446	0,02539 1,00000	77,506 22,494
Mf4	30	0,3764	0,15974 1,00000 20,22384	86,255 13,219 0,526	0,6487	0,36962 1,00000 4,00674	72,090 24,482 3,429	0,6267	0,06556 1,00000 4,59422	54,731 41,425 3,844
Mf5	39	0,3420	0,18579 1,00000 5,68643	88,263 10,437 1,294	0,2946	0,11012 1,00000 6,15938	85,049 13,955 0,996	0,3717	0,06106 1,00000 34,12811	84,678 14,819 0,503
Zaldia	22	0,2437	0,13794 1,00000	87,733 12,267	0,2857	0,14470 1,00000	83,509 16,491	0,2481	0,14761 1,00000	88,217 11,783
Mf1	9	0,2580	0,04784 1,00000	77,932 22,068	0,1640	0,12932 1,00000	96,019 3,981	0,2690	0,05969 1,00000	77,738 22,262
Mf2	7	0,3366	0,17812 1,00000	80,718 19,282	0,2685	0,08150 1,00000	79,640 20,360	0,2240	0,09710 1,00000	85,945 14,055
Mf3	5	0,4517	0,03234 1,00000	56,662 43,338	4,6829	0,07822 1,00000 46,13872	64,384 26,139 9,474	0,3976	0,01506 1,00000	61,161 38,839

Aztertutako saguaren ERVen gene eta metafamilia gehienetan hautespen ia-neutrala hautespen positiboa baino probableagoa izan zen (IV-8. taula). Hala ere, hautespen positibopean zeuden kodoiak aztertutako sekuentzia guztien hiru geneetan (*gag* kodoien %0,571ean, *pol* %0,201ean *env* %0,976an), Mf1-ko *env*en (%16,939), Mf4-ko *pol*en (%1,790), Mf6-ko *gag*en (%4,446) eta *pol*en (%2,374) eta *env*en (%2,685) detektatu ziren. Honetaz gain, kodoi gehienak hautespen purifikatzailepean egon ziren eta hautespen neutralpean pozioen %2,368tik %41,515era bitartean.

Arratoia

Aztertutako arratoiaren ERVen gene eta metafamilia gehienetan sekuentzian zehar 3 ω egotea bakarra baino probableagoa izan zen (IV-7. taula). Salbuespenak Mf3-ko *env* eta Mf5-eko *pol* eta *env* izan ziren. Mf1, Mf2, Mf3, Mf4 eta Mf5 metafamilietako *gag* genean izan ezik, kodoi gehienak hautespen purifikatzailepean egon ziren. Aipatutako metafamilietako *gag* geneetako kodoi gehienak, aldiz, hautespen positibopean egon ziren. Adibide hauetaz gain, Mf1-eko eta Mf2-ko *pol* genean hautespen positibopean zeuden kodoiak detektatu ziren (kodoien %0,822 eta %8,302 hurrenez hurren).

Hautatutako arratoiaren ERVen gene eta metafamilia gehienetan hautespen ia-neutrala hautespen positiboa baino probableagoa izan zen (IV-8. taula). Mf1, Mf2, Mf3, Mf4 eta Mf5 metafamilietako *gag* genean izan ezik, kodoi gehienak hautespen purifikatzailepean egon ziren. Metafamilia hauetan kodoi gehienak hautespen neutralpean egon ziren eta hautespen positibopean zeuden kodoiak Mf1-en (%32,528), Mf2-n (%7,006), Mf3-n (%1,578) eta Mf5-n (%9,321) detektatu ziren *gag* genean. Adibide hauetaz gain, Mf2-ko *pol* genean ere hautespen positiboa kodoien %0,780 detektatu zen. Gainontzeko gene eta metafamilietan hautespen neutrala kodoien %2,957tik %24,482ra bitartean detektatu zen.

Txakurra

Aztertutako txakurraren ERVen gene eta metafamilia guztietan sekuentzian zehar 3 ω ezberdin onartzen duen eredua ω bakarra baino probableagoa izan zen (IV-7. taula), Mf1 metafamiliako *env* genean izan ezik. Bestalde, sekuentzian zehar kodoi gehienak hautespen purifikatzailepean egon ziren. Aipatutako Mf1-eko *env* izan ezik, aztertutako gainontzeko gene eta metafamiliek hautespen positibopean zeuden kodoiak azaldu zituzten (kodoien %1,955tik %31,926ra bitartean).

Hautespen ia-neutral eta positiboaren artean (IV-8. taula), aztertutako ERV gene eta metafamilia gehienetan hautespen ia-neutrala probableagoa izan zen. Hala ere, aztertutako sekuentzia guztien *pol* eta *env* geneetan eta Mf2 eta Mf3-ko *env* genean hautespen positibopean zeuden kodoiak detektatu ziren (kodoien %0,820tik %4,836era bitartean). Honetaz gain, kodoi gehienak hautespen purifikatzailepean egon ziren, Mf1-eko *env* genean izan ezik (kodoien %39,342). Hautespen neutralpean, aldiz, kodoien %10,531etik %60,658ra bitartean egon ziren.

Behia

Sekuentzia osoan zehar ω bakarra egoteko aukera edo 3 ω ezberdin egotearen artean (IV-7. taula), behiaren ERV gene eta metafamilia guztietan, 3 ω ezberdin onartzen dituen eredua probableagoa izan zen. Kasu guztietan sekuentziaren kodoi gehienak hautespen purifikatzailepean egon ziren. Hala ere, gene eta metafamilia gehienetan hautespen purifikatzailean zeuden kodoiak detektatu ziren. Salbuespenak sekuentzia guztien *pol* genea, Mf1 metafamiliako *gag* eta *env* geneak eta Mf3-ko *pol* eta *env* izan ziren, zeintzuetan bakarrik hautespen purifikatzailea detektatu zen. Hautespen positibopean zeuden kodoien proportzioa aldakorra izan zen, kasu batzuetan oso txikia izanik (adibidez, sekuentzia guztien *gag* genean kodoien %0,512) eta beste batzuetan oso handia (Mf4 metafamiliako *env* genean kodoien %44,702).

Hautespen ia neutrala eta hautespen positiboaren artean eredu probableena aldakorra izan zen (IV-8. taula). Hala ere, sekuentziaren kodoi gehienetan hautespen purifikatzailea detektatu zen, Mf1 metafamiliaren *gag* genean izan ezik (kodoien %48,190-k $\omega < 1$ izan zuten). Hautatutako behiaren ERV guztien sekuentziak analizatzerakoan, hiru geneen kodoiek hautespen positiboa (kodoien %0,374tik %0,804ra bitartean) azaldu zuten eta hautespen neutralpean kodoien %6,136tik %12,815era bitartean egon ziren. Metafamiliak aztertzerakoan, Mf4 eta Mf5 metafamilien hiru geneek eta Mf2-n *gag*ek hautespen positiboa azaldu zuten (kodoien %0,503tik %5,389ra bitartean). Kasu hauetan hautespen neutralpean kodoien %10,437tik %41,425era bitartean egon ziren; aztertutako gainontzeko kasuetan $\omega = 1$ kodoien %5,725etik %51,810era bitartekoa izan zen.

Zaldia

Aztertutako zaldiaren ERVen gene eta metafamilia guztietan sekuentzian zehar 3 ω ezberdin onartzen duen eredua probableena izan zen (IV-7. taula). Bai gene zein metafamilia guztietan kodoi gehienek $\omega < 1$ azaldu zuten. Gainera hainbat kasutan hautespen positibopean zeuden kodoiak detektatu ziren: Mf1 metafamiliaren *gag* (kodoien %21,082) eta *pol* (%0,905) geneetan, Mf3-ko *pol*en (%10,602) eta *env* gene guztietan (sekuentzia guztietan kodoien %9,749, Mf1-en 3,349, Mf2-n %10,910 eta Mf3-n %19,896).

Hautatutako zaldiaren ERVen gene eta metafamilietan hautespen ia-neutrala hautespen positiboa baino probableagoa izan zen (IV-8. taula), Mf3 metafamiliako *pol* genean izan ezik. Kasu horretan kodoien %9,474ak $\omega > 1$ izan zuten eta $\omega = 1$ %26,139a. Gainontzeko kasuetan hautespen neutrala

kodoien %3,981tik %43,338ra bitartean egon zen. Hala ere, Mf3 metafamiliko *gag* eta *env* geneek gainontzeko metafamiliek baino kodoi gehiago azaldu zituzten hautespen neutralpean.

Laburpena

Oro har, aztertutako espezie eta metafamilia ia guztietan sekuentzien kodoi gehienetan hautespen purifikatzailea eman zen, $\omega < 1$ izan baitzen (IV-7. taula eta IV-8. taula). Emaitza honen arabera, badirudi gene erretrobiralek funtzioa mantentzeko presioa jasan dezaketela sekuentziaren proportzio handi batean.

Hala ere, espezie eta metafamilia gehienetan $\omega = 1$ edota $\omega > 1$ zuten kodoiak detektatu ziren, hau da, hautespenean ez zeuden eskualdeak edota eskualdeak zeintzuetan berrikuntza bultzatuta zegoen. Gainera, erabilitako bi analisiak, oro har, bat etorri ziren hautespen positibopean zeuden kodoien proportzioa detektatzerakoan.

Kasurik aipagarriena arratoiaren *gag* genea da, zeinetan metafamilia guztietan kodoi gehienetan $\omega = 1$ edo $\omega > 1$ izan baitzen, hots, hautespen purifikatzailepean zeuden kodoiak nahiko gutxi izan ziren.

IV.4. Eztabaida

Lan honetan 7 mamalioen genometik hautatutako erretrobirus endogenoen (ERV) errekonbinazio intergenikoaren, errekonbinazio intragenikoaren eta hautespenaren analisi sistematikoa burutu zen ERV hauen kopia-kopuruaren emendioa eta sekuentzietan baimendutako aldaketak, hots, dinamika ebolutiboa, aztertzeko asmoarekin.

Lehenik eta behin esan beharra dago hemen aurkeztutako emaitzak ERVen azpitalde baten emaitzen isla direla. Zehazki, ERV hauek 3 gene erretrobiral nagusiak nahiko ondo mantendu zituztenak izan ziren. Ezagarrri hau zuten ERV gehienak I klasekoak izan ziren. Hortaz, emaitza hauek ondo kontserbatutako ERVetara alboratuta egon daitezke eta batez ere I klasera zuzenduak. Harrigarriki, ondo kontserbatutako ERV hauek txertaketa hein zabala izan zuten: ERV batzuk nahiko zaharrak izan ziren eta beste batzuk, aldiz, nahiko berriak. Adibidez, behiaren Mf5 metafamiliak 0.2 eta 40 miloi urte arteko txertaketa heina azaldu zuen eta txinpantzearen Mf3-k 1 eta 53 miloi urte artekoa. Hortaz, ERVen adinean ez zen alborapenik egon.

Baita ere, analisisetan sakontzeko, ERVak metafamilia izeneko talde handitan taldekatu ziren eta metafamilia hauetako batzuetan ERV familia bat baino gehiagoko kideak taldekatu ziren. Horrela egitearen asmoa metafamiliak ahalik eta inklusiboenak egitea zen, lan honetan aztertu nahi ziren prozesu ebolutiboak, hotskopuru-emendioa eta sekuentzian baimendutako aldaketak, filogenetikoki gertu dauden ERVen artean gertatzea litekeena delako (Gifford & Tristem, 2003)

Bestalde, erabilitako metodologiek ere bere mugak izan zituzten. Errekonbinazio intergenikoa detektatzeko eraikitako zuhaitz filogenetikoek hainbat eragozpen izan zituzten. Alde batetik, hiru geneak batera egindako zuhaitzetan izan ezik, zuhaitz askok ez zuten nahi besteko bereizmenik eduki, adar batzuek ez baitzuten sendotasun estatistikorik izan. Gertakari honen jatorrian aztertutako nukleotido-luzeran eta ERVen dibergentzia-maila egon daitezke. Hortaz, historia ebolutibo ezberdina izan zuten ERVak eta euren geneak adierazterakoan soilik sendotasun estatistikoa izan zuten taldekapenak kontutan hartu ziren aztertutako espezieetan, hots, gizaki, txinpantze, sagu, arratoi, txakur, behi eta zaldian.

Errekonbinazio intrageniko aztertzerakoan bost metodo ezberdin erabili ziren eta metodo ezberdinek batzuetan bat egin zuten eta besteetan ez. Mota honetako analisiak egiteko metodo bat baino gehiago erabiltzea gomendagarria da eta, gainera, erabilitako metodoak oinarri ezberdina edukitzea (Posada, 2002). Hortaz, bi metodo filogenetiko (RDP eta PLATO), eta ordezkaren hiru metodo (Maxchi, Chimaera eta Geneconv, azken hau gene konbertsioa espresuki detektatzeko diseinatu dagoena) erabili ziren. Metodo bakoitzak errendimendu ezberdina du datuen arabera eta faltsu-positiboen eta faltsu-negatiboen tasa ezberdinak dituzte (Posada, 2002). Hori dela eta, alde batetik zuzenketarik zorrotzena erabili zen (Bonferroni zuzenketa) errekonbinazio seinaleak ahalik eta esangarrienak zirela ziurtatzeko eta, bestetik, metodo ezberdinen emaitzak uztartu ziren metodo

gehienek bat egiten zuten eskualdeak kontutan hartzeko.

Azkenik, metodologiari dagokionez, adierazi beharra dago gizakien ERVetan hautespena kalkulatzeko metodo ezberdinak erabili direla ERVek aurkezten dituzten lerrokatzeko arazoak direla eta (Belshaw et al., 2004; Belshaw et al., 2005). Programa horiek gure datuak lerrokatzerakoan huts egin zuten, hots, ez ziren aproposak gure datuak lerrokatzeko. Hortaz, hautespena kalkulatzeko modua guk diseinatutako metodologia erabili genuen eta, agian, kontserbazioaren gainestima egon daitezke irakurtarauak mantendu nahian.

Lan honetan gizakian dauden ERVak ere aztertu ziren, lehendik ere azterketa sakona jaso bazuten ere (Belshaw et al., 2004; Belshaw et al., 2005). Orain arte aipatutako eragozpenen eragina “neurtzeko” gizakiaren ERVak analizatzea baliagarria izan zen, eragozpenek eragin zitzaketen alborapenen “kontrol” modura eta, horrela, eragozpen hauek maneiatzeko. Bestalde, zazpi genometan detekzioetatik analisietara metodologia berdina erabiltzeak, genoma ezberdinak balizko alborapen berdinekin konparatzea ahalbidetu zuen.

Behin analisi hauen mugak argitu eta finkatu direla, emaitzak aztertuz espezie ezberdinetan hautatutako ERVek, euren dinamika ebolutiboan, hainbat joera komun izan zituzten. LTR eta zati zentralaren arteko erlazioari dagokionez, karraskarietan, oro har, LTR eta geneek jatorri berdina azaldu zuten. Behian eta gizakian, ordea, jatorri ezberdina zuten hainbat kasu detektatu ziren, hala nola, behiaren Mf5 metafamiliako hainbat kide edo gizakiaren Mf4-ko hainbat kide. Bi espezie hauetako hainbat ERVen LTR eta zati zentralaren jatorria ez zen berdina izan, lehendik gizakien ERVetan ezaguna zen gertaera (Belshaw et al., 2004; Hughes & Coffin, 2005). Berriki egindako ikerketa batean (Benachenhou et al., 2009), I klaseko ERVetan *pol* genearen eta LTRen zuhaitzek, oro har, adarkadura berdina zutela ondorioztatu zuten, II klasekoetan, aldiz, ezberdintasunen bat azaltzen zuten *pol* genearen eta LTRen zuhaitza erkatzerakoan (Benachenhou et al., 2009). Gertaera hauek, barne-zatien arteko errekonbinazioekin baino, solo-LTR, genomaz zehar zati zentralik gabe agertzen diren LTRak, eta ERV osoak baino askoz ere ohikoagoak diren egiturak, eta barne zatien arteko gene konbertsio edo errekonbinazioarekin erlazionatuta egon daitezkeela aditzera eman da (Belshaw et al., 2004).

Lan honetan, lehen aldiz, errekonbinazioa intragenikoa detektatzeko diseinatutako programak erabiliz, ERVetan egindako errekonbinazio intragenikoari buruzko azterketa espezifikoak aurkezten da. Orain arte prozesu hau gizakien (Belshaw et al., 2004; Hughes & Coffin, 2005; Kjeldbjerg et al., 2008) eta txerrien (Lee et al., 2008) ERVetan aztertu da soilik. Kasu batzuetan errekonbinazio intragenikoa analisi filogenetikoaren sendotasuna neurtzeko erabili izan da (Belshaw et al., 2004) edo ez ziren programa espezifikoak erabili (Hughes & Coffin, 2005; Lee et al., 2008). Salbuespena gizakiaren ERV-V familiaren *env* geneari buruzko lan batean dago (Kjeldbjerg et al., 2008), zeinean Geneconv programa erabili zen eta familia horren *env* gene konbertsioa detektatu zen. Bestalde, txerriaren *env* geneen artean ere errekonbinazioa detektatu zen (Lee et al., 2008), baina prozedura nahiko sinple baten bidez egindako azterketaren bidez.

Oro har, errekonbinazio intragenikoa espezie eta metafamilia bakoitzean modu ezberdinean gertatu zela ikusi zen. Hala ere, kasu batzutan, hala nola, behiaren ERVen hiru geneetan edota saguaren ERVen genetean, seinale errekonbinante sendoa zuten eskualdeak egon ziren. Hortaz, modu aldakorrean bada ere, errekonbinazioa gertatu gertatzen dela ondorioztatu daiteke. Salbuespen bakarra zaldiaren ERVak izan ziren, oso gertaera gutxi detektatu baitziren. Erabilitako ERV kopurua (22 sekuentzia) mugatua izan bazen ere, txakurraren pareko kopurua (23 sekuentzia) izan zen eta espezie honetako ERVetan errekonbinazio-gertaerak detektatu egin dira. Gainera, Posadak errekonbinazioa detektatzeko erabilitako sekuentzia-kopuruak eraginik ez zuela iradoki zuen (Posada, 2002) eta lan honetan, RDP metodoan izan ezik, ez dugu detektatutako errekonbinazio-gertaera eta sekuentzia-kopuruaren artean korrelaziorik ikusi. Hortaz, erabilitako ERV kopuruaren eragina baztertu daiteke eta zaldiaren genomaz ERVek dinamika ezberdina izan dezaketela ondorioztatu liteke.

Errekonbinazioa erretrobirus exogenoetan nahiko ohikoa den prozesu bat da (Robertson et al.,

1995; Vogt, 1997), hortaz, euren kide endogenoek gaitasun hori mantentzea ez litzateke horren harrigarria izango. Gainera, ERVen arteko homologia maila altua kontutan hartuta, errekonbinazio-gertaerak ematea ezin daiteke baztertu (Kjeldbjerg et al., 2008). Homologian oinarritutako errekonbinazioaz gain, erretrubirus endogenoen arteko errekonbinazio ez-homologoak egon daitezkeela aditzera eman da (Jern & Coffin, 2008). Ildo honekin jarraituz, saguetan ERVekin erlazioa duten IAP (*Intracistinal A-type particle*) elementuen eta erretrubirala ez den RNA-birus exogeno baten arteko errekonbinazio ez-legitimoa deskribatu da (Geuking et al., 2009). Adibide guzti hauek hemen aurkeztutako ERVen errekonbinazio intragenikoaren gertaera indartzen dute.

Hautespenaren eta errekonbinazio intergenikoaren emaitzak uztartuz ERVen kopuru emendioari buruz hainbat ondorio lor daitezke. Honetarako *env* genearen hautespena berrinfekzioa erretrubirubosiziotik ezberdintzeko gakoa da: dn/ds proportzioa (ω) 1 baino txikiagoa bada, funtzioa mantentzen duen heinean, berrinfekzioa gertatu dela ondoriozta daiteke. Bestalde, erretrubirubosizioaren barnean *cis* era eta *trans* era ezberdintzeko *gag* eta *pol* geneen ω erabiltzen da: bietan 1 baino txikiagoa bada, funtzioa mantentzen dutenez, kopuru emendioa *cis* eran gertatu dela ondoriozta daiteke. Azkenik, hiru geneen ω ez bada 1 baino txikiagoa, kopuru emendioa *trans* eraren bidez gertatu dela ondorioztatu daiteke (Belshaw et al., 2004; Belshaw et al., 2005; Katzourakis et al., 2005). Gainera, errekonbinazio intergenikoak ERV baten geneen jatorriari buruz informazioa ematen duenez, jatorri ezberdina duten geneei buruz informazioa ematen digu.

Hortaz, behietan eta zaldietan, bi informazio hauek batuz, ERV gehienak berrinfekzio bidez ugaritu direla ondoriozta daiteke, gizakian, txipantzean, saguetan eta txakurran gertatzen den era berean. Alde batetik, oro har, hiru gene erretrubiral nagusietan hautespena purifikatzailea izan zen, iadanik gizakiaren ERVen *pol* (Gifford & Tristem, 2003) eta *env* (Benit et al., 2001) geneetan ezaguna zena eta ERVen funtzioa mantentzearekin lotzen dena. Bestetik, oro har, ERV gutxiren geneek jatorri ezberdina izan zuten, hots, errekonbinazio intergenikorik gertatu ez zela eta, ondorioz, beste ERV batek utzitako generik ez zutela euren egituraren barneratu. Gainera, hautespenari dagokionez, purifikatzailea ez ezik, barneko adarrek kanpokoek baino ω txikiagoa azaldu zuten. Gertaera hau ERVen aitzindariak, hots ez finkatutako ERVak, hautespenean egotearekin erlazioa duela aditzera eman da eta, hau, *trans* erako eboluzioa oso ohikoa ez izanarekin lotu da (Belshaw et al., 2004). Adierazi beharra dago hautespen purifikatzailea deigarria izan daitekeela DNA zabortzat jotzen diren elementu hauetan. Hala ere, gorago aipatu bezala, gene erretrubiraletan hautespen purifikatzailearen deskribapena lehengokoa da (Benit et al., 2001; Gifford & Tristem, 2003) eta aurrerago ω balio hauen interpretazioaz arituko gara.

Hautespen purifikatzailearen orokortasun honetatik hainbat gene eta metafamilia urrundu ziren. Hainbat analisisan barneko adarrek $\omega > 1$ azaldu zuten, adibidez, behiaren Mf3 metafamiliako *env* geneak, zaldiaren Mf3 metafamiliako *gag*ek, txinpantzearen Mf6-ko *gag* eta *pol*ek eta saguaren Mf2-ko *pol* eta *env* eta Mf3-eko *pol*. Salbuespen hauetan, behiaren, zaldiaren eta arratoiaren salbuespenen kasuan (arratoiaren Mf4 metafamilian izan ezik), gene horien metafamilia horietan errekonbinazio intergeniko kasuak ere detektatu ziren. Beraz, kasu hauetan, erretrubirubosizioa gertatu izana ezin daiteke baztertu. Txinpantzearen Mf6 metafamiliaaren kasuan, ordea, *gag* eta *pol*en barneko adarrek $\omega > 1$ azaldu bazuten ere, historia ebolutibo ezberdina *env* geneak izan zuen. Bestalde, saguaren salbuespenen kasuan ez zen gene horien errekonbinazio intergenikorik detektatu. Salbuespen hauen ω ikusita, hots, gene bat behintzat hautespen purifikatzailepean ez zegoenez, berrinfekzioa baztertu behar da (Belshaw et al., 2005). *gag* eta *pol* $\omega < 1$ mantentzen ziren kasuetan kopuru emendioa *cis* erako erretrubirubosizio bidez gertatu zela inferitu daiteke (behiaren Mf3 metafamilia) eta bi gene hauetako batean $\omega > 1$ bazen, *trans* erako erretrubirubosizioaren bidez ugaritu zirela (zaldiaren Mf3 metafamilia, txinpantzearen Mf6, saguaren Mf2 eta Mf3). Analisi hauetan kasurik deigarriena arratoiaren metafamilia guztietako *gag* geneak izan ziren: 5 metafamilietan $\omega > 1$ izan zen adar guztietan eta metafamilia guztietan, Mf4 izan ezik, errekonbinazio intrageniko gertaerak detektatu ziren. Hortaz, emaitzek arratoian metafamilietan sailkatzen diren ERVen kopia-kopuruaren emendioa *trans*-erretrubirubosizioan gertatu zela aditzera ematen dute.

Adierazi beharra dago hemen aurkezten diren emaitzak berrinfekzioaren gainestima bat izan daitezkeela. Gorago aipatu bezala, gizakia kontrol moduan edukitzea emaitzak neurtzeko lagungarri izan daiteke. Gizakien ERVetan berrinfekzioa nagusi bada ere (Belshaw et al., 2004; Belshaw et al., 2005), familia batzuk transposizio bidez ugaritu dira (Belshaw et al., 2005): HERV-H familiaren klado bat *cis* eran ugaritu zen, HERV-W-ren heren bat LINE-en bidez eta ERV9 *trans* eran. HERV-H-ko ERVak lan honetako Mf4-ean barneratu ziren, ziur asko gehienak *cis* eran ugaritu ez diren kladokoak, *env* genea hautespenean zegoela detektatu baitzen, berrinfekzio bidez ugarituko kladoan gertatzen dena (Belshaw et al., 2005). HERV-W familiarekin antzeko gertaeraren aurrean egon gitezke: lan honetan Mf6 metafamilian barneratu zen eta ez zenez *env* genean hautespenaren erlaxaziorik ikusi, hemen analizatutako ERV gehienak familia horretako berrinfekzio bidez ugaritu diren ERVak izatea pentsa daiteke (Belshaw et al., 2005).

Gizakiaren ERV9 familiaren kasua korapilatsuagoa da. Familia hau ere Mf6 metafamilian barneratu zen eta metafamilia honetan ere *pol* jatorri ezberdina zuten ERVak multzo bat osatu zuten. Agian, hemen, hainbat seinale ezberdinak gurutzatu ziren. Alde batetik, lehenago aipatu bezala, ERV9 familian *pol* eta *env* geneetan hautespena erlaxatua zegoela aditzera eman da eta, hortaz, *trans* erako erretrotransposizioa ugaritze mekanismoa nagusi zela (Belshaw et al., 2005). Era berean, familia honetan egindako azterketa sakon batean (Costas & Naveira, 2000), familia honen eboluzioa paraleloan eboluzionatu duten azpifamiliataz osatuta zegoela eta aktibitatearen bat mantentzea zezakeela aditzera eman zen. Agian gure emaitzetan orain dela gutxi arte aktibo mantendu diren familia honen ERVak edota azpifamiliaren gainerrepresentazioa egon zen eta ez *trans* eran ugaritu ziren ERV edo azpifamilienak.

Emendio-mekanismoei aritzerakoan aipatu bezala hautespena, oro har, purifikatzailea da. Hala ere, $\omega < 1$ horren azpian hautespen purifikatzailepean dauden kodoietaz gain, metafamilia gehienetan hautespen neutraldu edota positibodun kodoiak proportzio ezberdinean detektatu ziren.

Lan honetan hautespena nolakoa den aztertzeko lau eredu testatu ziren (sekuentzia osoan zehar ω bakarra, sekuentzia osoan zehar 3 ω , hautespen ia-neutrala eta hautespen positiboa) eta binakako konparaketak egin ziren (ω bakarra vs 3 ω ; ia-neutrala vs positiboa) eredu probableena aukeratzeko. Oro har, hautespen positiboa detektatu zen kasuetan, sekuentzian zeharreko ω -etako bat bat baino handiagoa zen ere. Hala ere eredu probableena aukeratzeko erabilitako bi analisiak bat etorri ez ziren kasuak egon ziren. Ohiko kontraesan bat sekuentzian zeharreko hiru ω -etako bat, bat baino handiagoa izanik ere hautespen positiboa ia-neutrala baino probableagoa ez izatea izan zen. Kasu hauetan, askotan, $\omega > 1$ hori ez zen batetik asko aldentzea emaitzak faltsu positibotzat hartzea gomendatzen da eta ia-neutrala vs hautespen positiboko analisiaren emaitzak lehenestea (Yang, 1997 – PAMLen gidaliburua). B

Honekin lotuta, hautespen positibopean zeuden proportzioak txikiak izan zireneko kasuetan, errekonbinazioaren eragina ezin daiteke baztertu. Dirudienez, errekonbinazio-gertaera asko daudenean, egiantz-arrazoi test bidezko analisisien (lan honetan burututakoen) faltsu positiboen tasa handitzen da (Anisimova et al., 2003). Beraz, hautespen positiboa proportzio txikian detektatu direnean faltsu positiboak direla ezin daiteke baztertu.

Salbuespenak alde batera utzita, hautespen purifikatzailepean, neutralpean eta positibopean dauden kodoiak detektatu ziren. Gorago aipatu bezala, gizakiaren ERVen geneak hautespen purifikatzailepean daudela ezaguna bada ere (Benit et al., 2001; Gifford & Tristem, 2003; Belshaw et al., 2004; Belshaw et al., 2005), hautespen neutralaren eta positiboaren eragina ez da horren ezaguna. Gizakien eta txinpantzen ERV-K familian (lan honetan gizakiaren Mf2 eta txinpantzearen Mf4 metafamiliatan barneratzen direnak) hautespen neutrala eta positiboa detektatua izan da ere (Romano et al., 2006).

Alde batetik, hautespen neutrala DNA zaborran itxaron daiteken hautespena da. Adibidez, arratoiaren ERVen *gag* genean gertatzen dena. Hautespen-mota hau pseudogeneek nagusiki azaltzen duten hautespena den heinean (Torrents et al., 2003), ERVetan hautespen neutrala agertzeak pseudogeneek jasotzen duten prozesua jasotzen ari direla adieraz lezake. Gainera, sekuentzian zehar

$\omega < 1$ ezberdinak ematen dira, batzuk besteak baino altuagoak, hautespenaren erlaxazioarekin erlazionatuta egon daitezkenak. Hortaz, ezin daiteke baztertu, gene erretrobiralen inaktibazioa graduala izatea eta kodoi gehienek (baina ez guztiek) hautespen purifikatzailea mantentzen dutenez, sekuentzia osoaren $\omega < 1$ izatea.

Bestetik, hautespen positiboa detektatzea deigarriagoa da. Gorago azaldu bezala errekonbinazioak hautespen positiboaren gainestima eragiten du. Hala ere, badaude proportzio altutan hautespen positiboa azaldu duten adibideak: behiaren Mf2-ko *gag*ek (kodoien %5.39), zaldiaren Mf3-ko *polek* (%9.47) edo saguaren Mf5-eko *gag*ek (%5.63). XRVetan hautespen positiboa detektatu bada ere (Lemey et al., 2005), ERVetan hautespen-mota honen agerpena azaltzea zailagoa litzateke. XRVen kasuan hautespen positiboa ostalariaren erantzun immunea saihesteko estrategia izan daitekeela proposatu da (Lemey et al., 2005) baina ERVen kasuan azalpen hau ez litzateke guztiz argigarria izango, ostalariaren genomaren parte diren heinean. Esan bezala, hainbat kodoietan berrikuntza baimenduta egotea nahiko harrigarria da, batez ere DNA zabortzat hartu ohi diren elementuetan. Erretrobirus edo erretrobirusen pareko elementuetan halako emaitzak “domestikatu” diren gene erretrobiraletan, hots, ostalariak bere onerako erabiltzen dituen geneetan, ikusi da, hala nola, saguan *gag* gene batean jatorria duen Fv1 genean (Qi et al., 1998) eta *Drosophila* *Iris* elementuan (Malik & Henikoff, 2005). Lan hauetan espezie ezberdinetatik lortutako elementu homologoetan hautespenaren analisiak egin ziren eta sekuentziak edo zenbait kodoi hautespen positiboan zeudela ikusi zen. Elementu hauen kasuan, erretrobirusen aurreko defentsan lagundu dezaketenez, erretrobirus ezberdinak ezagutzeko presioa izatea eta, ondorioz, hautespen positiboan egotea proposatu da (Qi et al., 1998 ; Malik & Henikoff, 2005). Hala ere emaitza hauek mekanismoen analisi sakonagoa beharko lukete eta gertaera hau azaltzen dituen mekanismo ezberdinen proposamena.

Laburbilduz, lan honetan hainbat espezieren ERVen azpimultzo baten dinamika ebolutiboa aztertu zen. Lortutako emaitzak anitzak izan baziren ere, espezie ezberdinetan hainbat orokortasun ikusi ziren. ERVen kopuruaren ugaritzeari zegokiola, espezie gehienetan berrinfekzioaren bidez gertatu zen. Salbuespena arratoien ERVak izan ziren, zeintzuek *gag* gene ez-funtzionala azaldu zuten hautespenaren analisisan eta, hortaz, ugaritzeko *trans*-erretrotransposizioa erabili izana ezin daiteke baztertu. Sekuentziaren osotasunaren mantentzeari zegokiola, errekonbinazio intragenikoaren gertaerak ohikoak izan ziren, zaldiaren ERVetan izan ezik. Gainera sekuentzian zehar hautespen-patroi anitza azaldu zuten, sekuentziaren zati gehiena hautespen purifikatzailean baitzeuden, baina hautespen positiboa zuten zatiak ere detektatu ziren. Guzti honek ERV hauen dinamika ebolutibo konplexu eta bizi bat azaldu digu, ERVak mutazioak pairatzen dituen eta inaktiboak diren DNA-zaborraren sekuentzia-iruditik urrunduz.

IV.5. Bibliografia

- Altschul, S., T. Madden, A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D. Lipman. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* **25**:3389-3402.
- Anisimova, M., R. Nielsen, and Z. Yang. 2003. Effect of recombination on the accuracy of the likelihood method for detecting positive selection at amino acid sites. *Genetics* **164**:1229-1236.
- Baillie, G., L. van de Lagemaat, C. Baust, and D. Mager. 2004. Multiple groups of endogenous betaretroviruses in mice, rats, and other mammals. *Journal of Virology* **78**:5784-5798.
- Belshaw, R., A. Katzourakis, J. Paces, A. Burt, and M. Tristem. 2005. High copy number in human endogenous retrovirus families is associated with copying mechanisms in addition to reinfection. *Mol Biol Evol* **22**:814-817.
- Belshaw, R., V. Pereira, A. Katzourakis, G. Talbot, J. Paces, A. Burt, and M. Tristem. 2004. Long-term reinfection of the human genome by endogenous retroviruses. *Proceedings of the National Academy of Sciences of the United States of America* **101**:4894-4899.
- Benachenhou, F., V. Blikstad, and J. Blomberg. 2009. The phylogeny of orthoretroviral long terminal repeats (LTRs). *Gene* **448**:134-138.
- Benit, L., P. Dessen, and T. Heidmann. 2001. Identification, phylogeny, and evolution of retroviral elements based on their envelope genes. *Journal of Virology* **75**:11709-11719.
- Boeke, J. and J. Stoye. 1997. Retrotransposons, endogenous retroviruses, and the evolution of retroelements, p. 343-435. In J. Coffin, S. Hughes, and H. Varmos (eds.), *Retroviruses*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

- Castresana, J.** 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Biol Evol* **17**:540-552.
- Costas, J. and H. Naveira.** 2000. Evolutionary History of the Human Endogenous Retrovirus Family ERV9. *Mol Biol Evol* **17**:320-330.
- Geuking, M. B., J. Weber, M. Dewannieux, E. Gorelik, T. Heidmann, H. Hengartner, R. M. Zinkernagel, and L. Hangartner.** 2009. Recombination of Retrotransposon and Exogenous RNA Virus Results in Nonretroviral cDNA Integration. *Science* **323**:393-396.
- Gifford, R., P. Kabat, J. Martin, C. Lynch, and M. Tristem.** 2005. Evolution and distribution of class II-related endogenous retroviruses. *Journal of Virology* **79**:6478-6486.
- Gifford, R. and M. Tristem.** 2003. The evolution, distribution and diversity of endogenous retroviruses. *Virus Genes* **26**:291-316.
- Grassly, N. and E. Holmes.** 1997. A likelihood method for the detection of selection and recombination using nucleotide sequences. *Mol Biol Evol* **14**:239-247.
- Guindon, S. and O. Gascuel.** 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic Biology* **52**:696-704.
- Hughes, J. and J. Coffin.** 2005. Human endogenous retroviral elements as indicators of ectopic recombination events in the primate genome. *Genetics* **171**:1183-1194.
- Jern, P. and J. Coffin.** 2008. Effects of Retroviruses on Host Genome Function. *Annual Review of Genetics* **42**:709-732.
- Johnson, W. and J. Coffin.** 1999. Constructing primate phylogenies from ancient retrovirus sequences. *Proceedings of the National Academy of Sciences of the United States of America* **96**:10254-10260.
- Katoh, K., K. Kuma, H. Toh, and T. Miyata.** 2005. MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Research* **33**:511-518.
- Katzourakis, A., A. Rambaut, and O. Pybus.** 2005. The evolutionary dynamics of endogenous retroviruses. *Trends in Microbiology* **13**:463-468.
- Kjeldbjerg, A., P. Villesen, L. Aagaard, and F. Pedersen.** 2008. Gene conversion and purifying selection of a placenta-specific ERV-V envelope gene during simian evolution. *BMC Evolutionary Biology* **8**.
- Klymiuk, N., M. Muller, G. Brem, and B. Aigner.** 2006. Phylogeny, recombination and expression of porcine endogenous retrovirus {gamma}2 nucleotide sequences. *J Gen Virol* **87**:977-986.
- Lee, D., J. Lee, N. Park, Y. Oh, M. Kwon, and Y. Kim.** 2008. Analysis of natural recombination in porcine endogenous retrovirus envelope genes. *Journal of Microbiology and Biotechnology* **18**:585-590.
- Lemey, P., S. Van Dooren, and A. Vandamme.** 2005. Evolutionary dynamics of human retroviruses investigated through full-genome scanning. *Mol Biol Evol* **22**:942-951.
- Malik H.S., and S. Henikoff.** 2005. Positive Selection of Iris, a Retroviral Envelope-Derived Host Gene in *Drosophila melanogaster*. *PLOS Genetics*, **1**: e44.
- Martin, D. and E. Rybicki.** 2000. RDP: detection of recombination amongst aligned sequences. *Bioinformatics* **16**:562-563.
- Martin, D., C. Williamson, and D. Posada.** 2005. RDP2: recombination detection and analysis from sequence alignments. *Bioinformatics* **21**:260-262.
- McCarthy, E. and J. McDonald.** 2003. LTR_STRUC: a novel search and identification program for LTR retrotransposons. *Bioinformatics* **19**:362-367.
- Posada, D.** 2002. Evaluation of methods for detecting recombination from DNA sequences: Empirical data. *Mol Biol Evol* **19**:708-717.
- Posada, D. and K. Crandall.** 1998. MODELTEST: testing the model of DNA substitution. *Bioinformatics* **14**:817-818.
- Posada, D. and K. Crandall.** 2001. Evaluation of methods for detecting recombination from DNA sequences: Computer simulations. *Proceedings of the National Academy of Sciences of the United States of America* **98**:13757-13762.
- Qi C.F., F. Bonhomme, A. Buckler-White, C. Buckler, A. Orth, M.R. Lander, S.K. Chattopadhyay, H.C. Morse III.** 1998. Molecular phylogeny of Fv1. *Mammalian Genome* **9**: 1049-1055.
- R Development Core Team.** 2008. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
- Robertson, D., P. Sharp, F. McCutchan, and B. Hahn.** 1995. Recombination in HIV-1. *Nature* **374**:124-126.
- Romano, C., R. Ramalho, and P. Zanutto.** 2006. Tempo and mode of ERV-K evolution in human and chimpanzee genomes. *Archives of Virology* **151**:2215-2228.
- Ronquist, F. and J. Huelsenbeck.** 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**:1572-1574.
- Sawyer, S.A. A computer package for the statistical detection of gene conversion. Distributed by the author, Department of Mathematics, Washington University in St. Louis . 1999.
- Ref Type: Electronic Citation
- Smith, J.** 1992. Analyzing the mosaic structure of genes. *Journal of Molecular Evolution* **34**:126-129.

Mamalian endogenous retrovirus (ERV) dynamics

Torrents, D., M. Suyama, E. Zdobnov, and P. Bork. 2003. A Genome-Wide Survey of Human Pseudogenes. *Genome Research* **13**:2559-2567.

Vogt, V. 1997. Retroviral virions and genomes, p. 343-435. *In* J. Coffin, S. Hughes, and H. Varmos (eds.), *Retroviruses*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Yang, Z. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. *Computer Applications in the Biosciences* **13**:555-556.

IV.6. Eranskina

IV-1. Taula gehigarria: Espezie bakoitzean erabilitako ERVak.

ERVa	Kromosoma:hasi-bukatu	ERVa	Kromosoma:hasi-bukatu	ERVa	Kromosoma:hasi-bukatu
Gizakia		chr5_131	5: 170358958-170367111 (-)	chr13_45	13: 84340941-84350950 (+)
chr1_7	1: 9069586-9080461 (+)	chr6_12	6: 16564177-16574467 (+)	chr13_48	13: 86206356-86217062 (-)
chr1_10	1: 12690538-12698398 (-)	chr6_14	6: 19379728-19388548 (+)	chr14_2	14: 20238541-20246289 (+)
chr1_41	1: 52453915-52464693 (-)	chr6_27	6: 31403045-31413715 (-)	chr14_9	14: 21775303-21783255 (+)
chr1_43	1: 55148928-55157051 (-)	chr6_35	6: 38692469-38701938 (+)	chr14_15	14: 31243399-31251054 (-)
chr1_47	1: 58406208-58415012 (+)	chr6_47	6: 63844178-63854816 (-)	chr14_33	14: 52664896-52673720 (+)
chr1_51	1: 66654097-66664512 (-)	chr6_48	6: 64093170-64101245 (-)	chr14_35	14: 57656828-57664426 (+)
chr1_72	1: 87038877-87047577 (-)	chr6_49	6: 64717744-64726145 (+)	chr14_39	14: 69828659-69838808 (-)
chr1_76	1: 94608750-94617109 (-)	chr6_59	6: 68646352-68657006 (-)	chr14_47	14: 82586116-82596100 (+)
chr1_81	1: 102322786-102333153 (-)	chr6_63	6: 74670809-74678400 (+)	chr15_24	15: 53382994-53393205 (-)
chr1_90	1: 105453477-105461675 (+)	chr6_72	6: 85477200-85486244 (-)	chr15_30	15: 80451565-80461811 (-)
chr1_105	1: 152910899-152919101 (+)	chr6_76	6: 89179253-89186755 (-)	chr15_34	15: 80953337-80963589 (+)
chr1_112	1: 155508449-155517119 (-)	chr6_82	6: 99814877-99823562 (-)	chr16_13	16: 21547619-21556434 (-)
chr1_116	1: 159487548-159495939 (+)	chr6_100	6: 119045591-119056109 (+)	chr17_8	17: 23580378-23594091 (-)
chr1_117	1: 159709290-159717433 (+)	chr6_108	6: 123103008-123112807 (-)	chr17_9	17: 25335267-25344205 (-)
chr1_122	1: 167898182-167906259 (+)	chr6_138	6: 165389764-165399910 (-)	chr17_11	17: 27113901-27124614 (+)
chr1_124	1: 169540925-169550283 (+)	chr6_140	6: 167225629-167233544 (+)	chr17_17	17: 38761989-38779060 (*)
chr1_126	1: 171603922-171619664 (+)	chr7_2	7: 4593456-4604941 (-)	chr17_23	17: 48005825-48015248 (+)
chr1_146	1: 197326320-197335359 (+)	chr7_6	7: 69554567-6966050 (-)	chr18_25	18: 38295492-38305417 (+)
chr1_153	1: 208427235-208437397 (+)	chr7_13	7: 12593724-12602561 (-)	chr18_34	18: 51768360-51777991 (+)
chr1_155	1: 210094336-210100408 (-)	chr7_19	7: 21552203-21562675 (-)	chr19_1	19: 5108163-5117120 (-)
chr1_167	1: 223139331-223149538 (+)	chr7_25	7: 29641235-29651888 (-)	chr19_4	19: 8350645-8359435 (+)
chr1_170	1: 225087549-225094423 (+)	chr7_36	7: 46466207-46476732 (-)	chr19_9	19: 9558055-9569143 (-)
chr2_5	2: 8951919-8960454 (-)	chr7_39	7: 52366525-52375014 (-)	chr19_24	19: 20058250-20074641 (-)
chr2_27	2: 41458141-41468061 (+)	chr7_40	7: 53988841-53998031 (-)	chr19_27	19: 20334800-20345448 (+)
chr2_31	2: 47181605-47189887 (+)	chr7_42	7: 64086935-64098071 (-)	chr19_30	19: 20487419-20498851 (-)
chr2_32	2: 47662122-47672880 (-)	chr7_51	7: 88006422-88015237 (-)	chr19_32	19: 20720175-20730930 (-)
chr2_62	2: 87119381-87125918 (+)	chr7_53	7: 91934243-91953235 (-)	chr19_48	19: 23151785-23164515 (*)
chr2_63	2: 87229593-87236130 (-)	chr7_59	7: 99158391-99167134 (+)	chr19_68	19: 39734131-39743184 (-)
chr2_70	2: 97804282-97814216 (+)	chr7_62	7: 106256947-106267897 (+)	chr19_80	19: 51296150-51305357 (+)
chr2_115	2: 166271090-166281432 (-)	chr7_70	7: 113805004-113812537 (-)	chr19_89	19: 57815382-57828623 (+)
chr2_122	2: 173085974-173098943 (*)	chr7_86	7: 133884684-133894950 (+)	chr20_4	20: 7202977-731221 (+)
chr2_158	2: 231662734-231674175 (-)	chr7_87	7: 134509310-134519001 (+)	chr20_28	20: 42731981-42740493 (+)
chr2_159	2: 231973447-231983543 (+)	chr7_104	7: 151354982-151372880 (-)	chr21_4	21: 16685052-16693733 (-)
chr2_161	2: 238777268-238787423 (-)	chr8_4	8: 5666119-5676655 (-)	chr22_7	22: 37781301-37789646 (-)
chr3_16	3: 16030740-16041132 (-)	chr8_15	8: 8022521-8030749 (+)	chrX_1	X: 3895193-3906012 (+)
chr3_18	3: 16780995-16790646 (+)	chr8_22	8: 9697824-9705988 (+)	chrX_10	X: 25968436-25977400 (+)
chr3_20	3: 19497091-19507880 (-)	chr8_28	8: 12385648-12396432 (-)	chrX_17	X: 34724101-34732540 (+)
chr3_32	3: 33463429-33471919 (+)	chr8_36	8: 27404599-27413996 (-)	chrX_44	X: 62217136-62227162 (+)
chr3_43	3: 38454296-38463130 (-)	chr8_41	8: 35909370-35918045 (+)	chrX_51	X: 64264600-64272449 (+)
chr3_48	3: 44558838-44567557 (-)	chr8_47	8: 39851592-39862337 (-)	chrX_52	X: 64496284-64504397 (+)
chr3_51	3: 46126334-46135995 (-)	chr8_50	8: 43524296-43534939 (-)	chrX_56	X: 72392215-72402805 (-)
chr3_54	3: 46402928-46410439 (-)	chr8_63	8: 50498650-50512266 (+)	chrX_72	X: 79099265-79107061 (+)
chr3_85	3: 86667645-86676469 (+)	chr8_65	8: 52363975-52374540 (-)	chrX_79	X: 82148182-82157208 (-)
chr3_87	3: 87701183-87709364 (+)	chr8_66	8: 54030059-54038480 (+)	chrX_91	X: 92571103-92581690 (+)
chr3_88	3: 87737679-87748486 (-)	chr8_67	8: 54145093-54152376 (-)	chrX_99	X: 96982499-96992938 (+)
chr3_91	3: 96788914-96797182 (-)	chr8_68	8: 54210307-54218822 (-)	chrX_122	X: 111237973-111246835 (-)
chr3_99	3: 105484389-105495742 (-)	chr8_70	8: 54534998-54545705 (+)	chrX_139	X: 117902510-117911464 (+)
chr3_120	3: 131101907-131110450 (+)	chr8_77	8: 61493159-61501967 (+)	chrX_142	X: 119164560-119176067 (+)
chr3_122	3: 131341133-131351930 (-)	chr8_85	8: 73463047-73471847 (+)	chrX_173	X: 150473087-150481497 (-)
chr3_147	3: 152900883-152909500 (+)	chr8_86	8: 74981106-74990524 (+)	chrX_174	X: 152012322-152021241 (-)
chr3_155	3: 158778561-158789198 (-)	chr8_98	8: 88005829-88016012 (-)	chrY_24	Y: 7919228-7927147 (-)
chr3_169	3: 165216768-165227510 (-)	chr8_105	8: 93473815-93482449 (-)	chrY_26	Y: 8128986-8140607 (-)
chr3_179	3: 168466720-168477235 (+)	chr8_109	8: 108677817-108688317 (-)	chrY_32	Y: 8908546-8918598 (+)
chr3_192	3: 177705113-177716011 (-)	chr8_137	8: 146214311-146230538 (*)	chrY_36	Y: 9173134-9185866 (-)
chr3_208	3: 192860221-192873151 (*)	chr9_11	9: 21912879-21923279 (+)	chrY_64	Y: 14779812-14790811 (+)
chr4_7	4: 4094226-4102279 (-)	chr9_12	9: 22424843-22433002 (+)	chrY_70	Y: 15127896-15136053 (+)
chr4_32	4: 31968661-31975985 (+)	chr9_33	9: 71728730-71738934 (-)	chrY_72	Y: 15433772-15442763 (-)
chr4_35	4: 35535764-35543385 (-)	chr10_11	10: 17092224-17100997 (-)	chrY_77	Y: 15882329-15890536 (+)
chr4_38	4: 44684981-44695085 (+)	chr10_15	10: 22468041-22476991 (+)	chrY_78	Y: 16006591-16015255 (-)
chr4_48	4: 59019299-59028011 (+)	chr10_23	10: 37603102-37613712 (-)	chrY_82	Y: 16692832-16703507 (+)
chr4_51	4: 63601458-63612205 (+)	chr10_24	10: 38264631-38275115 (+)	chrY_89	Y: 17256240-17266791 (-)
chr4_52	4: 63733240-63742910 (+)	chr10_70	10: 96512559-96523268 (-)	chrY_95	Y: 17805712-17814068 (-)
chr4_56	4: 68578922-68587645 (+)	chr10_86	10: 110791309-110799727 (+)	chrY_117	Y: 19753395-19763722 (+)
chr4_77	4: 77182316-77190429 (-)	chr10_92	10: 121633326-121641952 (+)	chrY_120	Y: 20039786-20055624 (+)
chr4_88	4: 89480496-89490108 (*)	chr11_18	11: 17348874-17358993 (+)	chrY_138	Y: 21657191-21665883 (-)
chr4_91	4: 92227869-92238658 (-)	chr11_22	11: 22969484-22979040 (-)	chrY_159	Y: 24380629-24390836 (+)
chr4_108	4: 102770026-102778612 (-)	chr11_26	11: 2391700-23919811 (+)		
chr4_110	4: 103375778-103387203 (+)	chr11_39	11: 39712081-39720856 (+)	Txinapantzea	
chr4_113	4: 108520431-108530851 (+)	chr11_49	11: 51208924-51218455 (-)	chr1_4	1: 12917013-12924875 (-)
chr4_143	4: 137094261-137102948 (-)	chr11_65	11: 67144378-67152488 (+)	chr1_9	1: 24162607-24173150 (+)
chr4_144	4: 137750707-137761143 (+)	chr11_72	11: 68148111-68163522 (+)	chr1_26	1: 53076921-53087873 (-)
chr4_162	4: 159798882-159807572 (+)	chr11_85	11: 76936342-76945914 (+)	chr1_28	1: 55894155-55902937 (-)
chr4_176	4: 171346532-171354509 (-)	chr11_94	11: 84995486-85004305 (+)	chr1_29	1: 59224141-59232968 (+)
chr4_186	4: 184156450-184164358 (+)	chr11_113	11: 100606121-100616761 (-)	chr1_42	1: 88247380-88256086 (-)
chr4_187	4: 185892882-185901329 (+)	chr11_118	11: 105597837-105608250 (+)	chr1_43	1: 95867368-95875726 (-)
chr4_188	4: 187398776-187406484 (-)	chr11_122	11: 109775730-109784565 (+)	chr1_53	1: 106791464-106799657 (+)
chr5_4	5: 7107025-7117577 (-)	chr12_3	12: 7225661-7232210 (+)	chr1_70	1: 140544558-140553196 (+)
chr5_21	5: 29425863-29433954 (+)	chr12_11	12: 8832432-8840555 (+)	chr1_74	1: 149064363-149072428 (+)
chr5_22	5: 30233789-30242667 (+)	chr12_32	12: 18075783-18086397 (+)	chr1_77	1: 152867598-152877017 (+)
chr5_24	5: 32587475-32595883 (-)	chr12_37	12: 20989020-20997916 (-)	chr1_98	1: 191682871-191691333 (+)
chr5_32	5: 42147473-42155239 (+)	chr12_61	12: 49581602-49591851 (-)	chr1_99	1: 192335648-192341722 (-)
chr5_39	5: 49702230-49710389 (+)	chr12_95	12: 91304806-91313805 (+)	chr1_100	1: 193512252-193522983 (-)
chr5_47	5: 62894916-62903919 (+)	chr12_109	12: 126202238-126210217 (+)	chr2a_3	2a: 9074532-9083662 (-)
chr5_57	5: 76239379-76247986 (+)	chr12_111	12: 126661320-126670343 (+)	chr2a_8	2a: 31991994-32002439 (+)
chr5_81	5: 111283043-111292117 (-)	chr12_112	12: 130385651-130396100 (+)	chr2a_18	2a: 48222874-48231156 (+)
chr5_85	5: 112981386-112989557 (-)	chr12_114	12: 131975913-131986830 (+)	chr2a_28	2a: 71964374-71975105 (-)
chr5_88	5: 114354632-114364917 (+)	chr13_8	13: 34615239-34621851 (-)	chr2a_35	2a: 80892995-80900937 (+)
chr5_112	5: 138863409-138873183 (+)	chr13_35	13: 66810970-66821487 (-)	chr2a_45	2a: 98708657-98718588 (+)
chr5_127	5: 165486430-165496862 (-)	chr13_36	13: 66869054-66879808 (+)	chr3_10	3: 16399556-16409894 (-)
				chr3_22	3: 34236022-34244499 (+)

Mamaloien erretrobirus endogenoen (ERV) dinamika ebolutiboa

ERVa	Kromosoma:hasi-bukatu	ERVa	Kromosoma:hasi-bukatu	ERVa	Kromosoma:hasi-bukatu
chr3_27	3: 37168724-37179394 (-)	chr16_9	16: 21875628-21884427 (-)	chr10_34	10: 33395533-33403811 (+)
chr3_32	3: 39303463-39312242 (-)	chr17_5	17: 14242563-14260175 (+)	chr10_35	10: 33435893-33445800 (-)
chr3_35	3: 45522496-45531238 (-)	chr17_8	17: 25487246-25498468 (-)	chr10_55	10: 41156345-41167006 (+)
chr3_38	3: 47154246-47163904 (-)	chr17_10	17: 27313665-27322275 (+)	chr10_83	10: 53208491-53218804 (-)
chr3_44	3: 52974426-52983085 (+)	chr17_11	17: 29078643-29086216 (+)	chr10_103	10: 78879495-78890233 (*)
chr3_66	3: 89918181-89928939 (-)	chr17_12	17: 29125296-29135052 (*)	chr10_107	10: 82844455-82855671 (-)
chr3_71	3: 99392776-99401025 (-)	chr17_24	17: 74460700-74469952 (*)	chr10_133	10: 117682660-117690358 (-)
chr3_80	3: 108324246-108335598 (-)	chr18_21	18: 46698939-46706829 (-)	chr11_11	11: 9780866-9791506 (-)
chr3_112	3: 156597527-156606273 (-)	chr18_24	18: 52353713-52363649 (+)	chr11_51	11: 27644463-27654922 (-)
chr3_113	3: 156635549-156644164 (+)	chr19_1	19: 5252161-5260982 (-)	chr11_63	11: 29949180-29959787 (-)
chr3_119	3: 162590545-162601178 (-)	chr19_3	19: 8613905-8622774 (+)	chr11_75	11: 50275793-50283456 (-)
chr3_133	3: 172431944-172442229 (+)	chr19_7	19: 9821104-9832225 (-)	chr11_81	11: 53846899-53856778 (+)
chr3_138	3: 181854470-181865263 (-)	chr19_19	19: 21465468-21476689 (-)	chr11_99	11: 86014168-86024766 (+)
chr4_15	4: 25370133-25378564 (-)	chr19_27	19: 2387367-23402027 (*)	chr11_101	11: 88750357-88758828 (-)
chr4_20	4: 36104653-36114747 (-)	chr20_15	20: 42005594-42013435 (+)	chr12_1	12: 3013224-3023594 (-)
chr4_57	4: 78985927-78996572 (+)	chr21_4	21: 16491242-16499924 (-)	chr12_22	12: 19464650-19471532 (-)
chr4_68	4: 94066957-94078002 (-)	chrX_22	X: 52966403-52977911 (-)	chr12_23	12: 19541369-19551752 (-)
chr4_77	4: 104924127-104932734 (-)	chrX_28	X: 62302294-62312641 (+)	chr12_29	12: 20230783-20241398 (-)
chr4_97	4: 136455767-136463698 (+)	chrX_33	X: 72585145-72595724 (-)	chr12_35	12: 21649070-21659706 (+)
chr4_101	4: 139440522-139448547 (+)	chrX_51	X: 92637399-92648418 (+)	chr12_46	12: 23311321-23321705 (-)
chr4_102	4: 139631035-139639735 (+)	chrX_67	X: 111617855-111626782 (-)	chr12_56	12: 24710353-24720681 (+)
chr4_106	4: 145470307-145478487 (-)	chrX_73	X: 118372480-118381974 (+)	chr12_59	12: 28567592-28577319 (+)
chr4_107	4: 145724580-145733110 (-)	chrY_2	Y: 553747-564361 (+)	chr12_91	12: 82524716-82535040 (+)
chr4_112	4: 162915454-162924146 (+)	chrY_9	Y: 1404276-1414515 (-)	chr12_101	12: 89329546-89341570 (+)
chr4_114	4: 163878944-163889726 (-)	chrY_12	Y: 1532410-1540828 (+)	chr13_7	13: 4331012-4341982 (*)
chr4_125	4: 187645517-187653972 (+)	chrY_14	Y: 1625506-1634673 (+)	chr13_12	13: 7622136-7632059 (-)
chr4_126	4: 189405354-189413800 (+)	chrY_32	Y: 5059250-5069488 (-)	chr13_22	13: 15489053-15499576 (-)
chr5_8	5: 38931120-38939720 (-)	chrY_44	Y: 7466138-7476726 (-)	chr13_37	13: 27537155-27545903 (-)
chr5_23	5: 65639693-65647847 (-)	chrY_74	Y: 12609142-12618309 (-)	chr13_38	13: 27556738-27567134 (-)
chr5_29	5: 73444476-73452238 (-)	chrY_102	Y: 16033028-16048857 (*)	chr13_44	13: 33530730-33541738 (*)
chr5_37	5: 83193454-83201409 (+)	chrY_110	Y: 16643042-16657545 (+)	chr13_45	13: 33644507-33654524 (+)
chr5_39	5: 85611962-85620834 (-)	chrY_111	Y: 16807423-16815785 (+)	chr13_75	13: 62541203-62551255 (+)
chr5_61	5: 113162299-113171926 (-)	chrY_127	Y: 17700507-17717278 (-)	chr13_79	13: 64929913-64937583 (-)
chr5_65	5: 114884320-114892497 (-)	chrY_144	Y: 19379221-19390216 (+)	chr13_89	13: 68120489-68134160 (+)
chr5_74	5: 127752827-127763155 (+)	chrY_153	Y: 20116579-20124786 (-)	chr13_92	13: 68252231-68260220 (+)
chr5_78	5: 141242562-141252075 (+)	chrY_159	Y: 20850237-20858115 (-)	chr13_116	13: 77822853-77833344 (-)
chr6_8	6: 19634395-19643229 (+)	chrY_167	Y: 21660623-21681990 (-)	chr14_9	14: 7490401-7504509 (+)
chr6_9	6: 23590359-23600328 (+)	chrY_170	Y: 21907198-21928564 (+)	chr14_46	14: 27148877-27157185 (+)
chr6_18	6: 31978082-31988789 (-)	chrY_184	Y: 23509095-23519537 (+)	chr14_92	14: 88813788-88823484 (-)
chr6_25	6: 39484721-39494190 (+)	Sagua		chr15_34	15: 34793163-34800403 (-)
chr6_40	6: 74735399-74742990 (+)	chr1_45	1: 26955186-26964161 (+)	chr15_62	15: 45743411-45753376 (+)
chr6_54	6: 100864438-100873150 (-)	chr1_48	1: 27786042-27796722 (+)	chr16_4	16: 3328085-3338999 (+)
chr6_73	6: 124848530-124858329 (-)	chr1_50	1: 30138697-30149232 (-)	chr16_7	16: 3462319-3473990 (+)
chr7_5	7: 6982218-6992804 (-)	chr1_82	1: 78492295-78502907 (-)	chr16_24	16: 12529467-12540350 (+)
chr7_13	7: 21833406-21843908 (-)	chr1_85	1: 83976269-83986854 (-)	chr16_47	16: 21023041-21032801 (+)
chr7_16	7: 30010885-30021069 (-)	chr1_86	1: 85132124-85142805 (-)	chr17_80	17: 46482525-46495908 (+)
chr7_24	7: 47232052-47242536 (-)	chr1_103	1: 104698179-104708761 (+)	chr17_83	17: 48316623-48329314 (+)
chr7_30	7: 54970901-54980107 (-)	chr1_139	1: 133539086-133549720 (+)	chr17_105	17: 61612575-61623314 (+)
chr7_35	7: 64636219-64647350 (-)	chr1_147	1: 142860270-142870935 (+)	chr18_13	18: 6415675-6423899 (-)
chr7_40	7: 88153425-88162563 (-)	chr1_168	1: 171945697-171955419 (+)	chr18_14	18: 6545619-6553259 (+)
chr7_56	7: 120371075-120381790 (-)	chr1_169	1: 172871394-172882052 (+)	chr18_22	18: 9136000-9146241 (-)
chr7_63	7: 135708845-135718511 (+)	chr2_34	2: 30082767-30093569 (+)	chr18_36	18: 17476693-17486705 (-)
chr8_10	8: 7717974-7726135 (+)	chr2_41	2: 36654819-36661768 (-)	chr18_43	18: 19007394-19017980 (+)
chr8_11	8: 7758398-7769192 (+)	chr2_51	2: 47385378-47392488 (+)	chrX_1	X: 3007933-3018618 (-)
chr8_26	8: 23963198-23972593 (+)	chr2_58	2: 57065686-57076347 (-)	chrX_48	X: 17476502-17484131 (-)
chr8_32	8: 32659958-32669127 (+)	chr2_60	2: 61699511-61707993 (-)	chrX_173	X: 33259699-33270302 (+)
chr8_37	8: 36586202-36595269 (-)	chr2_95	2: 128883909-128892211 (-)	chrX_234	X: 62307991-62315687 (-)
chr8_52	8: 49315991-49325871 (-)	chr2_100	2: 138342418-138353044 (+)	chrX_269	X: 85548547-85561462 (-)
chr8_54	8: 51148725-51155994 (-)	chr2_142	2: 181653872-181664536 (+)	chrY_2	Y: 193030-207235 (-)
chr8_55	8: 51214320-51222840 (-)	chr3_26	3: 11084202-11095383 (-)	Arratoia	
chr8_60	8: 51834214-51842224 (+)	chr3_67	3: 36018890-36031803 (*)	chr1_10	1: 2237072-2247757 (-)
chr8_65	8: 58533620-58541725 (-)	chr3_113	3: 94480260-94490884 (+)	chr1_21	1: 12804644-12815206 (+)
chr8_72	8: 70777922-70786681 (+)	chr3_137	3: 129746247-129754152 (+)	chr1_39	1: 31081490-31092171 (+)
chr8_91	8: 106735993-106746505 (-)	chr4_2	4: 3281470-3292070 (-)	chr1_47	1: 33891195-33901819 (+)
chr8_98	8: 137567087-137577419 (-)	chr4_13	4: 11101946-11112279 (+)	chr1_56	1: 58012377-58022930 (-)
chr9_2	9: 6194097-6204511 (+)	chr4_14	4: 12105810-12115270 (+)	chr1_74	1: 67207911-67218516 (+)
chr9_10	9: 22934031-22942197 (+)	chr4_15	4: 15162259-15170459 (+)	chr1_99	1: 82108511-82119443 (+)
chr9_12	9: 27417593-27425906 (+)	chr4_24	4: 25442529-25450836 (-)	chr1_173	1: 161813311-161823912 (-)
chr10_9	10: 17649512-17658306 (-)	chr4_47	4: 50350331-50360542 (-)	chr1_188	1: 201107246-201120286 (-)
chr10_11	10: 22683285-22692249 (+)	chr4_116	4: 112039591-112047247 (+)	chr1_212	1: 230713199-230721166 (+)
chr10_16	10: 38797151-38807977 (+)	chr4_149	4: 145684952-145695600 (-)	chr2_41	2: 23868768-23876829 (-)
chr10_48	10: 94992511-95003235 (-)	chr4_153	4: 146592702-146603351 (+)	chr2_47	2: 31249169-31259695 (+)
chr10_60	10: 109626876-109635292 (+)	chr5_1	5: 3019573-3029950 (-)	chr2_52	2: 49055843-49066333 (+)
chr10_64	10: 120653458-120662075 (+)	chr5_6	5: 6483309-6493993 (+)	chr2_98	2: 122152163-122162689 (+)
chr11_6	11: 13465594-13472636 (+)	chr5_20	5: 15939135-15950182 (+)	chr2_101	2: 128852452-128863015 (-)
chr11_8	11: 17319127-17329235 (+)	chr5_39	5: 24170483-24186129 (+)	chr2_113	2: 152582505-152593064 (-)
chr11_10	11: 23026053-23035902 (-)	chr5_70	5: 72953917-72964549 (+)	chr3_79	3: 96071835-96082857 (-)
chr11_53	11: 75965529-75974524 (+)	chr5_105	5: 105349529-105360201 (+)	chr3_97	3: 140969784-140980309 (+)
chr11_64	11: 84219494-84228312 (+)	chr5_137	5: 148365859-148373533 (-)	chr3_109	3: 161731206-161741992 (-)
chr11_78	11: 104885128-104895541 (+)	chr6_27	6: 36676875-36687070 (+)	chr4_7	4: 3819196-3826981 (*)
chr11_81	11: 109147532-109155862 (+)	chr6_43	6: 49444833-49452383 (+)	chr4_26	4: 13605270-13615405 (-)
chr12_5	12: 7488632-7495173 (+)	chr7_12	7: 6570458-6590049 (+)	chr4_127	4: 164358747-164369054 (+)
chr12_28	12: 30233128-30243575 (-)	chr7_23	7: 12185501-12193115 (+)	chr5_2	5: 36403-46928 (+)
chr12_32	12: 38752218-38762460 (+)	chr7_56	7: 18073760-18084319 (+)	chr5_33	5: 16727597-16738175 (+)
chr12_56	12: 68107846-68116742 (+)	chr7_62	7: 18964687-18975231 (+)	chr5_109	5: 106441308-106451867 (+)
chr12_75	12: 93297134-93305779 (+)	chr7_80	7: 21982757-21993423 (+)	chr5_116	5: 110217927-110227325 (+)
chr12_90	12: 129572936-129581961 (+)	chr7_97	7: 24091132-24101871 (-)	chr5_145	5: 164264388-164276772 (-)
chr12_91	12: 133407629-133418159 (+)	chr7_120	7: 28686354-28695054 (-)	chr6_14	6: 13978030-13989541 (*)
chr13_9	13: 34933485-34940082 (-)	chr7_141	7: 34810960-34821993 (+)	chr6_23	6: 24030227-24039162 (+)
chr13_16	13: 67384775-67395653 (-)	chr8_55	8: 69866637-69874482 (-)	chr6_54	6: 94939588-94950256 (+)
chr13_17	13: 76443171-76453746 (+)	chr8_72	8: 75922210-75930703 (+)	chr6_78	6: 129318248-129334734 (*)
chr13_23	13: 85347201-85357253 (+)	chr8_108	8: 125948387-125959126 (-)	chr7_31	7: 9403558-9413928 (-)
chr14_3	14: 19657629-19666342 (-)	chr9_45	9: 14965024-14975427 (+)	chr7_41	7: 13138037-13148620 (-)
chr14_8	14: 21175171-21183520 (+)	chr9_56	9: 19029974-19037661 (+)	chr7_62	7: 20788751-20802377 (+)
chr14_29	14: 82992960-83003076 (+)	chr9_118	9: 109492910-109500637 (-)	chr7_101	7: 74282449-74293157 (+)
chr15_11	15: 52715328-52725979 (-)	chr9_121	9: 109658327-109666159 (-)	chr7_112	7: 83488727-83499488 (+)

ERVa	Kromosoma:hasi-bukatu
chr8_3	8: 758970-769556 (+)
chr8_12	8: 4697014-4707257 (-)
chr8_35	8: 16476264-16487685 (+)
chr9_72	9: 113107913-113118689 (-)
chr10_4	10: 787798-797294 (+)
chr10_9	10: 4205101-4215868 (-)
chr12_3	12: 133633-144310 (+)
chr13_8	13: 3284085-3294599 (-)
chr14_9	14: 4363574-4372864 (+)
chr14_14	14: 8648836-8658229 (+)
chr14_53	14: 79892584-79903325 (-)
chr14_73	14: 103781546-103792107 (+)
chr15_83	15: 91460142-91470683 (-)
chr17_16	17: 20006868-20017632 (-)
chr18_9	18: 6054238-6064303 (+)
chr18_62	18: 71156051-71166562 (+)
chr20_2	20: 2611952-2622683 (-)
chr20_10	20: 4484587-4493774 (-)
chrX_8	X: 9203744-9214276 (+)
chrX_84	X: 65951924-65962486 (+)
chrX_97	X: 75102679-75113282 (+)
chrX_112	X: 91910799-91920480 (*)
chrX_151	X: 115313337-115323336 (+)
Txakurra	
chr1_4	1: 3322557-3330902 (-)
chr1_10	1: 51740444-51749181 (-)
chr1_18	1: 103013204-103023888 (+)
chr3_7	3: 41683161-41692133 (+)
chr3_10	3: 85014327-85024511 (-)
chr4_6	4: 38358514-38366784 (-)
chr7_7	7: 76563209-76572227 (+)
chr8_3	8: 11046411-11055891 (+)
chr8_12	8: 59645034-59651968 (-)
chr10_1	10: 6264697-6274907 (+)
chr11_7	11: 38701418-38710230 (+)
chr15_1	15: 14119857-14128320 (-)
chr15_2	15: 32846213-32854636 (+)
chr17_4	17: 49933256-49941441 (+)
chr20_4	20: 36645763-36659022 (-)
chr24_3	24: 7879541-7886715 (-)
chr25_4	25: 36496796-36506784 (+)
chr26_2	26: 31950327-31959215 (-)
chr26_3	26: 34325299-34335759 (+)
chr35_2	35: 27284269-27294140 (+)
chr35_4	35: 28040409-28051244 (-)
chrX_5	X: 43312682-43320810 (+)
chrX_6	X: 43822155-43831236 (+)
Behia	
1_5	1: 8625962-8635985 (-)
1_23	1: 76864086-76874394 (+)
1_34	1: 109771581-109780162 (+)
1_43	1: 119642767-119651367 (+)
1_52	1: 135862148-135870733 (+)
1_54	1: 137728046-137736386 (-)
1_58	1: 145500205-145507889 (+)
2_5	2: 1856883-1864414 (-)
2_6	2: 4046523-4055117 (+)
2_8	2: 6865738-6877980 (+)
2_15	2: 26960400-26968977 (-)
2_26	2: 69350995-69361048 (-)
2_34	2: 88708191-88719610 (+)
2_35	2: 89644805-89653575 (-)
3_5	3: 7448153-7457355 (+)
3_8	3: 9911997-9920968 (-)
3_13	3: 14947098-14958371 (+)
3_22	3: 27178385-27188347 (+)
3_44	3: 97256295-97264809 (+)
3_47	3: 110191444-110200042 (+)
4_4	4: 4646655-4656128 (-)
4_7	4: 5161199-5169814 (+)
4_18	4: 27408868-27418035 (-)
4_19	4: 28611868-28622158 (+)
4_20	4: 33488089-33501364 (+)
4_24	4: 40874131-40883012 (+)
4_26	4: 45855666-45866206 (-)
4_41	4: 98567342-98583598 (-)
4_44	4: 104979058-104989110 (+)
5_15	5: 40720694-40729627 (-)
5_26	5: 58139244-58149091 (+)
5_39	5: 100600441-100610458 (+)
6_14	6: 14658519-14667134 (+)
6_20	6: 26969323-26976989 (+)
6_22	6: 37492803-37502044 (-)
6_46	6: 102487500-102497561 (-)
7_6	7: 7977590-7986217 (+)
8_17	8: 42699574-42708193 (-)
8_29	8: 70005866-70015922 (-)
9_14	9: 26746755-26757441 (-)
9_32	9: 69574930-69582265 (-)
10_9	10: 21892543-21901517 (+)
11_12	11: 17083506-17091659 (-)
11_18	11: 32522822-32530490 (-)
11_31	11: 99683892-99691690 (+)
12_20	12: 37293618-37302919 (+)
12_31	12: 56017095-56026005 (-)
13_3	13: 22659349-22668215 (-)

ERVa	Kromosoma:hasi-bukatu
13_19	13: 73184078-73197136 (+)
13_22	13: 77896629-77919633 (-)
14_3	14: 22670025-22684435 (-)
14_7	14: 39011102-39019899 (-)
15_3	15: 8414194-8422163 (-)
15_7	15: 16150462-16160524 (-)
15_8	15: 16253468-16264348 (-)
15_18	15: 49302007-49327037 (+)
15_19	15: 49527014-49551862 (+)
15_23	15: 73269853-73279895 (+)
16_7	16: 5193188-5202190 (+)
16_19	16: 34953459-34961067 (-)
16_21	16: 42819360-42830121 (*)
17_4	17: 4795060-4804035 (-)
18_6	18: 14734098-14749262 (-)
18_18	18: 52092475-52101086 (-)
18_28	18: 56550392-56560439 (-)
19_17	19: 34885943-34896788 (-)
19_19	19: 36194623-36204816 (+)
19_27	19: 56940310-56953304 (-)
20_1	20: 8659053-8667636 (-)
20_15	20: 64131019-64139495 (-)
21_18	21: 32980977-32988658 (-)
23_14	23: 27514449-27524308 (+)
23_16	23: 27654628-27664824 (+)
23_18	23: 28157049-28166863 (-)
24_13	24: 47897308-47905807 (-)
24_16	24: 54784761-54792198 (-)
25_8	25: 15049437-15056495 (+)
26_5	26: 8210107-8221177 (+)
26_7	26: 9163919-9172972 (+)
26_18	26: 36537863-36547093 (-)
28_7	28: 11681840-11691910 (-)
29_1	29: 2145420-2153099 (+)
28_27	28: 39654356-39663304 (+)
29_3	29: 21070351-21079392 (+)
X_1	X: 3304565-3315196 (+)
X_7	X: 24850350-24863470 (+)
X_10	X: 27283530-27291179 (-)
X_21	X: 37547260-37557493 (-)
X_25	X: 46324322-46335252 (*)
X_40	X: 54904883-54913819 (-)
X_52	X: 63057274-63065872 (-)
X_57	X: 65103350-65111911 (-)
X_58	X: 65241297-65252481 (-)
Zaldia	
chr1_10	1: 46862874-46873226 (-)
chr2_4	2: 11712828-11722965 (+)
chr5_5	5: 16158775-16170731 (-)
chr5_7	5: 21511175-21520591 (+)
chr5_8	5: 27324361-27334495 (-)
chr5_14	5: 44862512-44872584 (-)
chr5_15	5: 77378682-77388710 (+)
chr7_12	7: 43692788-43702913 (-)
chr7_15	7: 46973867-46984109 (+)
chr7_18	7: 57912315-57919321 (-)
chr10_3	10: 6533264-6543822 (-)
chr10_7	10: 12944449-12951974 (+)
chr11_12	11: 55534814-55544496 (-)
chr15_12	15: 44643816-44654015 (-)
chr17_4	17: 37524978-37534774 (-)
chr20_9	20: 32761886-32771446 (-)
chr25_7	25: 35311451-35321373 (+)
chr27_3	27: 16321740-16329841 (+)
chrX_12	X: 44328148-44338555 (+)
chrX_15	X: 49902859-49912741 (+)
chrX_18	X: 53124020-53135308 (+)
chrX_24	X: 68875624-68883127 (+)

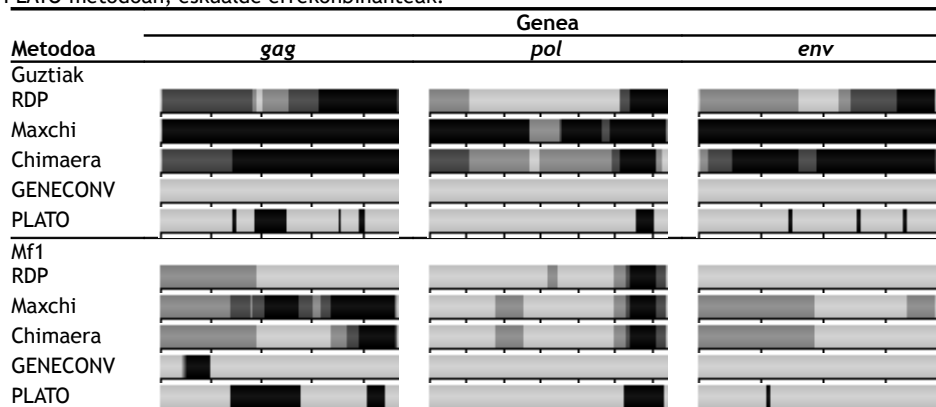
IV-2. Taula gehigarria: Hautespenean testatutako ereduen artean probableena aukeratzeko erabilitako egiantz maximoko test-aren p-balioak.

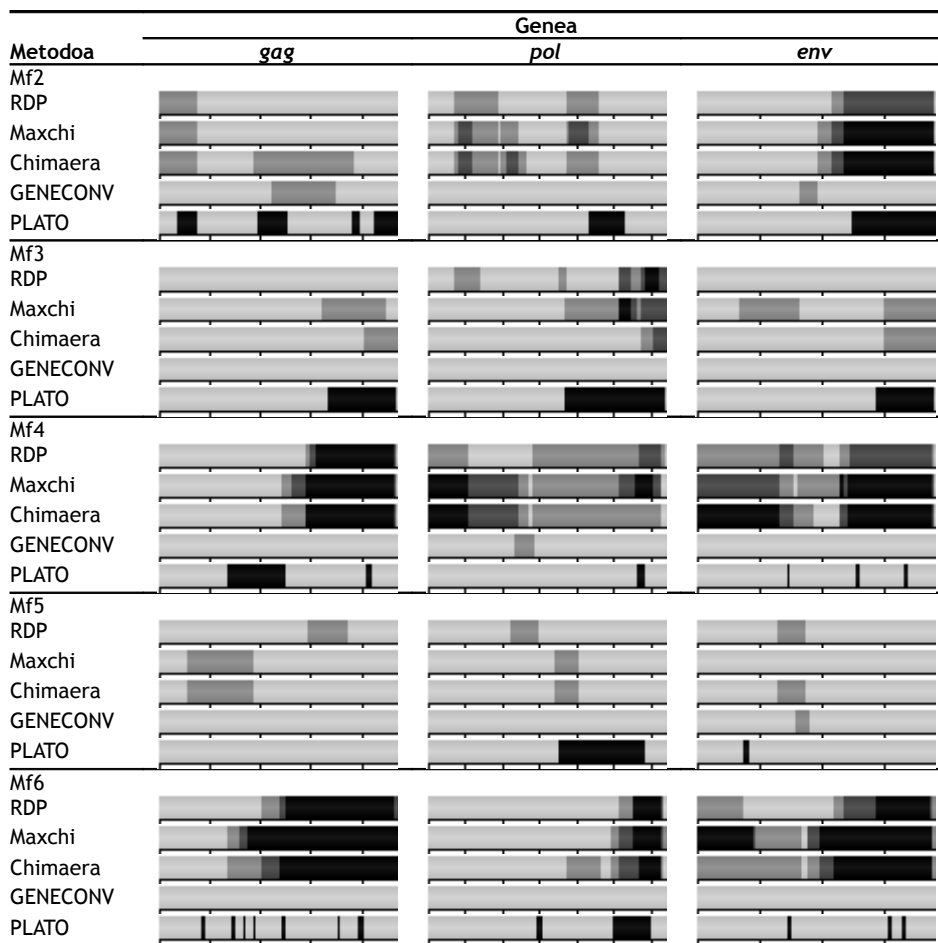
p<0,05 bada eredu konplexuena (Adar, M3, M2) probableagoa da. N/A, testak huts egin zuen eredu konplexuaren egiantz balioa eredu sinpleagoaren balioa baino txikiagoa izan baitzen.

Espezia	Genea								
	gag			pol			env		
	Adar vs M0	M3 vs M0	M2 vs M1	Adar vs M0	M3 vs M0	M2 vs M1	Adar vs M0	M3 vs M0	M2 vs M1
Gizakia	3,5973E-62	4,612E-236	1,4811E-34	7,697E-108	1,114E-251	1,9657E-14	3,9886E-21	4,902E-100	4,9013E-06
Mf1	0,00055472	1,2593E-76	4,145E-09	2,0446E-18	3,503E-100	2,661E-12	0,08001664	1,2934E-21	0,07388274
Mf2	3,5885E-11	2,4959E-29	0,06045718	6,3567E-19	4,2867E-18	1	9,4507E-07	3,074E-08	1
Mf3	0,22384665	8,7659E-18	0,01351214	1,4913E-18	9,4496E-32	1	0,00526247	0,09453938	0,999994
Mf4	4,1949E-09	1,5688E-18	1	2,2034E-15	4,3794E-34	1	0,00135837	3,7018E-15	0,82235426
Mf5	9,6733E-23	1,683E-181	6,5193E-24	1,2221E-30	8,785E-151	5,2044E-16	2,0466E-12	7,375E-100	0,01241751
Mf6	N/A	0,14657539	0,74292291	N/A	4,3302E-14	1,239E-07	#iNUM!	0,56254152	0,95176892
Txinpantzea	9,7343E-53	9,316E-163	1,2776E-24	6,5103E-39	4,043E-150	3,6519E-10	3,1688E-21	5,5395E-64	0,21734029
Mf1	4,3583E-16	1,0336E-15	1	7,1063E-07	7,1021E-07	1	0,00010844	1,0995E-06	1
Mf2	0,05664729	4,8701E-20	0,00127024	0,15984018	6,3629E-23	0,7031894	0,00188242	1,0257E-15	0,02876025
Mf3	9,6595E-07	5,7509E-30	1	0,13381149	2,8998E-55	1	4,9662E-07	6,4167E-20	1
Mf4	0,46174228	0,02358157	1	0,30018886	2,3899E-08	1	0,12534977	0,38894665	0,83912047
Mf5	0,08906016	5,1384E-06	0,38190821	0,21619902	4,7295E-16	7,2974E-05	0,56239637	0,48763548	N/A
Mf6	1,6754E-19	1,722E-159	1,073E-32	5,7676E-15	6,2734E-89	7,6752E-07	0,0009159	1,3936E-57	1
Mf7	0,0231997	0,01243636	1	3,4918E-05	0,2431154	1	0,42000902	0,00996404	1
Sagua	2,6842E-30	9,1305E-73	0,00361295	1,9189E-59	1,66E-153	2,5176E-07	0,00104401	2,7191E-62	0,00131277
Mf1	8,6281E-07	1,3169E-21	0,00247875	1,9298E-11	7,8372E-13	1	0,15867824	8,2785E-09	1
Mf2	0,00027216	1,9423E-12	0,00293971	1,0803E-07	3,6227E-47	0,00742774	0,03230002	5,0983E-08	1
Mf3	1,14E-09	3,8457E-18	1	6,5236E-11	1,0707E-35	1	1,7031E-09	0,13151349	1
Mf4	0,7842614	0,00884403	0,37848684	0,15060886	1,9839E-07	0,00932278	0,70482212	2,5079E-07	0,01175375
Mf5	0,43722371	1,1174E-06	N/A	9,434E-07	2,1334E-21	0,999999	3,9114E-05	N/A	N/A
Mf6	2,7388E-06	0,19028428	0,999989	6,5852E-05	1,3126E-11	0,00315564	0,01816264	0,00121473	0,55400642
Mf7	0,02517791	2,3749E-07	0,62955567	0,11315512	8,8693E-11	0,80732094	0,00949543	3,8556E-22	0,00311813
Arratoia	1,61E-12	3,9129E-55	0,999984	1,3954E-08	7,08E-79	1	0,00037269	1,0096E-42	N/A
Mf1	0,1383371	0,0010136	0,00333461	0,81520408	0,47499706	1	N/A	0,5043292	N/A
Mf2	0,22381598	0,01358291	0,3337864	0,10632876	0,01150018	1	0,97327884	0,04235111	1
Mf3	0,65777472	2,8119E-07	1,3282E-07	0,0167358	0,00509187	1	0,63280213	0,1754855	1
Mf4	0,26404291	1,387E-19	9,4674E-13	0,0049271	1,9062E-19	0,00971797	0,02763851	2,1225E-07	1
Mf5	0,16956629	7,8583E-08	3,414E-09	0,00823358	5,191E-16	0,2223519	0,00141815	0,01690678	1
Txakurra	0,21100447	2,7976E-18	0,999999	0,05205141	2,2001E-15	0,02080879	0,50804771	1,1021E-23	2,559E-11
Mf1	2,0004E-11	0,00027599	1	0,81122729	5,3654E-13	0,52979137	0,86119418	1,3821E-09	0,00137091
Mf2	0,07620074	0,00098146	0,1377351	0,86675197	0,00351047	0,99892158	0,00567484	0,20273017	0,70557444
Mf3	N/A	1,3506E-07	1	0,64679249	0,0008267	1	1,407E-199	9,0879E-13	4,3732E-07
Behia	7,8799E-13	7,9427E-94	3,1208E-27	2,98E-50	9,4379E-82	1,0255E-19	0,00230923	1,7212E-93	3,4731E-17
Mf1	0,08554023	9,8508E-33	3,771E-11	0,00017617	2,3439E-44	2,0136E-15	0,3353988	3,638E-54	0,00011244
Mf2	0,05471652	3,7834E-08	1	4,9309E-09	3,6022E-23	#iNUM!	0,0083098	0,01271271	1
Mf3	1,1073E-11	3,7827E-21	9,5051E-07	4,7235E-24	2,1332E-09	0,91017718	0,2721534	3,7465E-06	0,49076593
Mf4	0,64189989	4,6138E-14	1	0,00034492	5,5163E-11	1	0,00215852	0,00296139	0,55108622
Mf5	0,01306012	1,9315E-42	2,2256E-12	0,22300335	1,1004E-48	4,6167E-10	0,32455602	4,3952E-46	3,2087E-08
Zaldia	7,4037E-43	2,5872E-49	N/A	7,3528E-17	2,9302E-32	1	6,4649E-05	1,6743E-08	1
Mf1	7,3253E-11	4,7733E-08	1	0,77952957	8,8028E-06	1	0,00167647	0,00023546	1
Mf2	0,00057329	1,4389E-09	0,8893228	0,13457396	0,00242979	0,05623044	0,19407314	1,1605E-07	1
Mf3	0,02997074	1,1667E-17	0,90226591	0,52049619	7,8607E-11	0,00010211	0,01032023	5,1459E-06	0,66166295

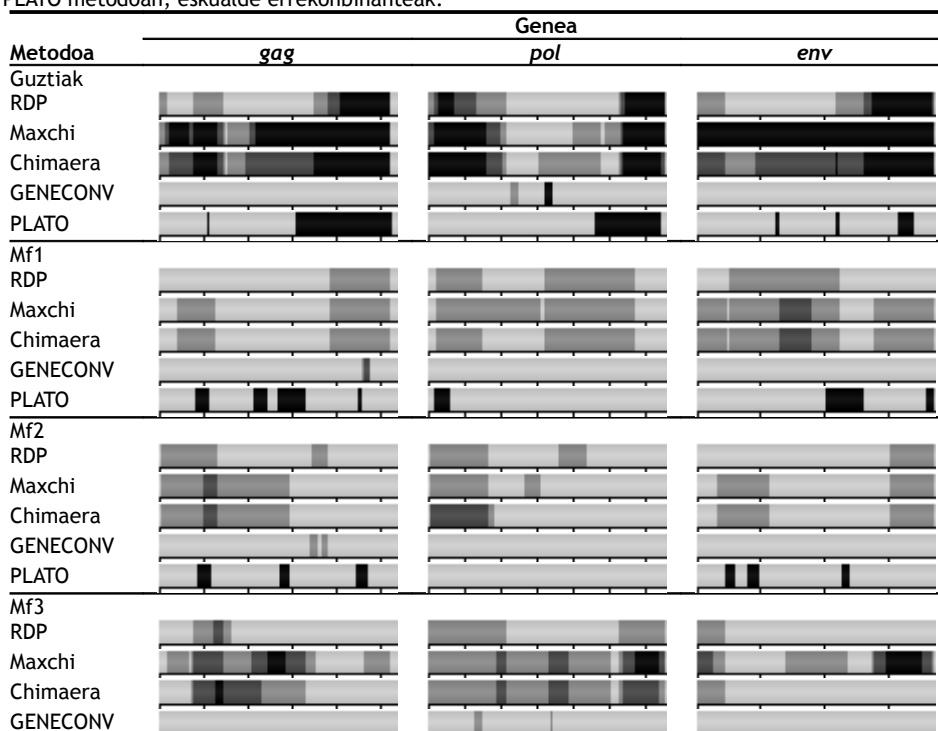
IV-3. Taula gehigarria: Aztertutako gizakiaren ERVen errekonbinazio intragenikoaren detekzioa.

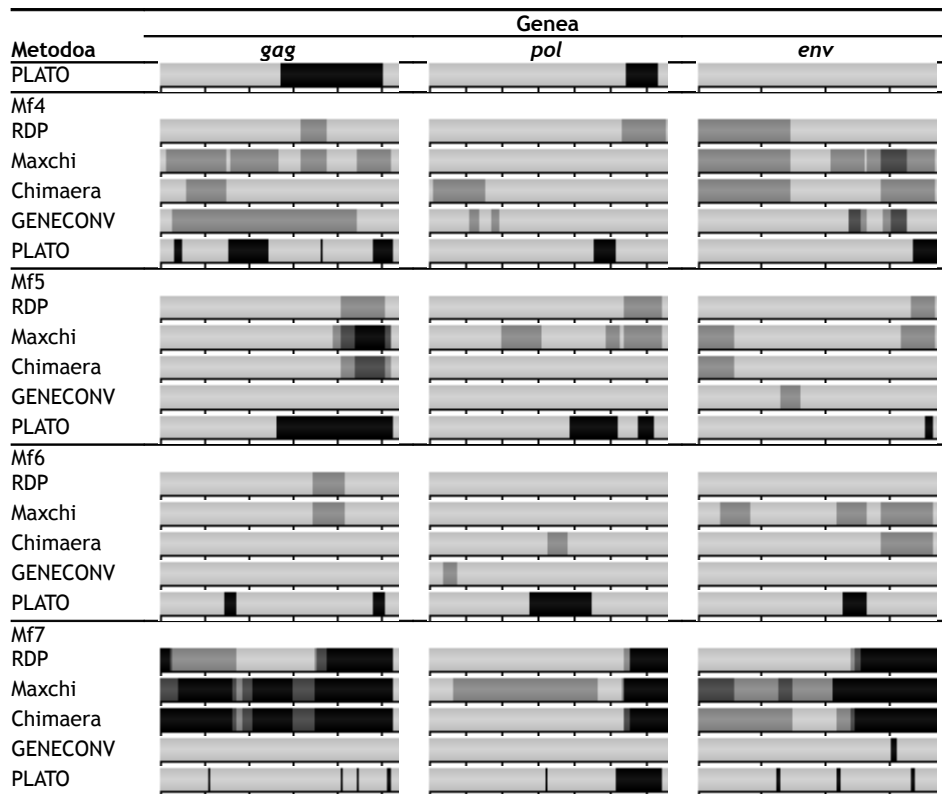
RDP, Maxchi, Chimaera eta GENECONV metodoetan beltz kolorearen intentsitateak errekonbinazio-gertaeren kopurua adierazten du. PLATO metodoan, eskualde errekonbinanteak.





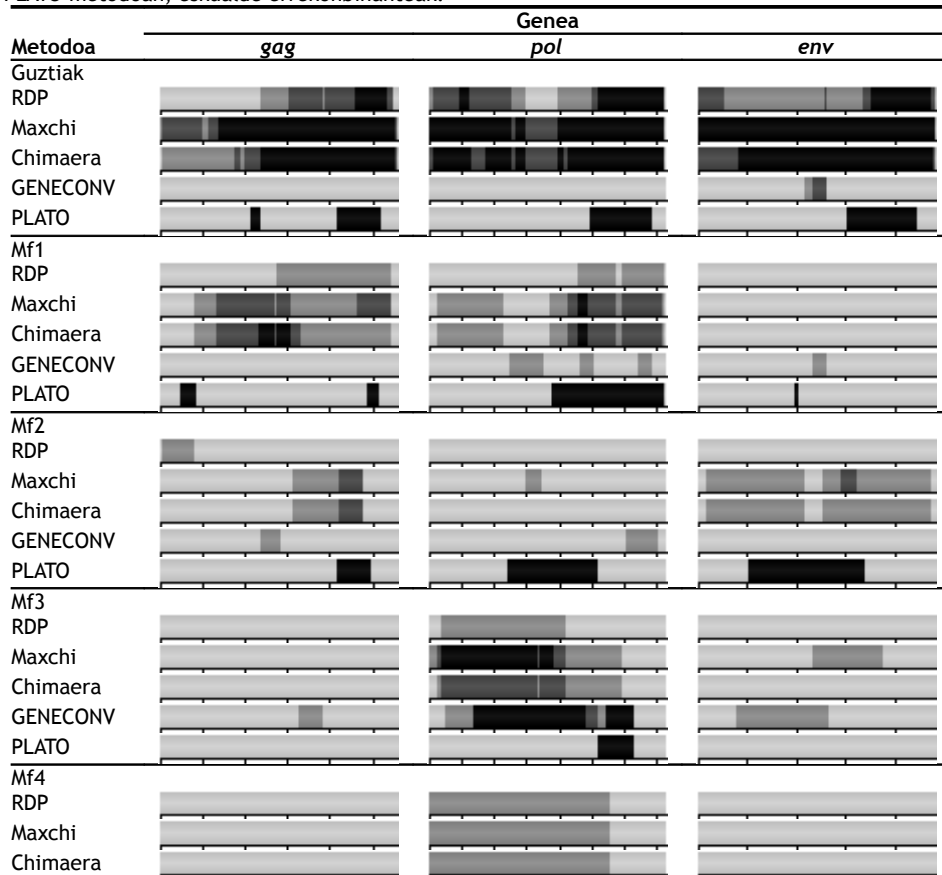
IV-4. Taula gehigarria: Aztertutako txinpantzearen ERVen errekonbinazio intragenikoaren detekzioa. RDP, Maxchi, Chimaera eta GENECONV metodoetan beltz kolorearen intentsitateak errekonbinazio-gertaeren kopurua adierazten du. PLATO metodoan, eskualde errekonbinanteak.

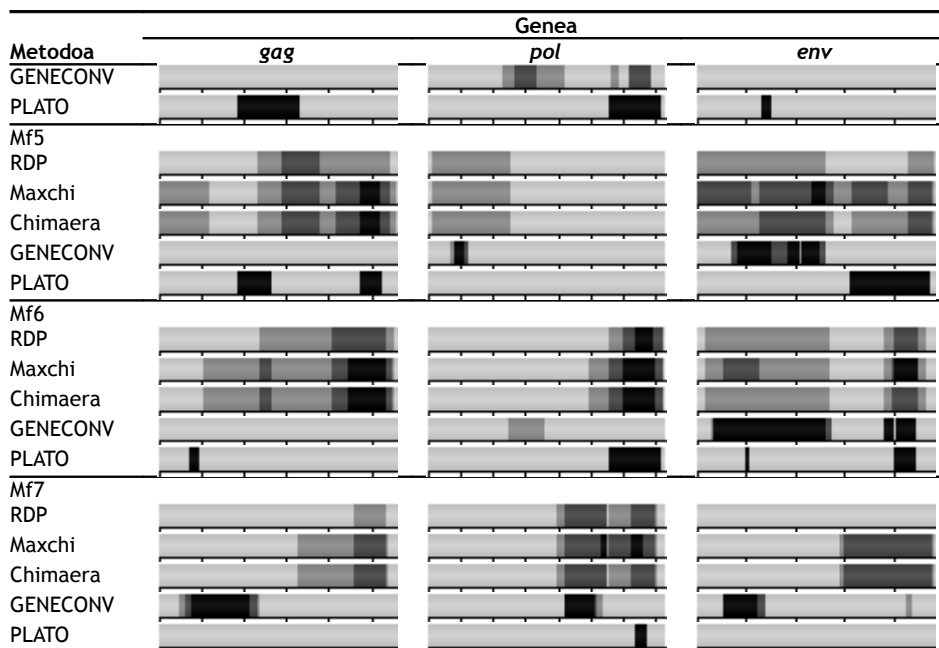




IV-5. Taula gehigarria: Aztertutako saguaren ERVen errekonbinazio intragenikoaren detekzioa.

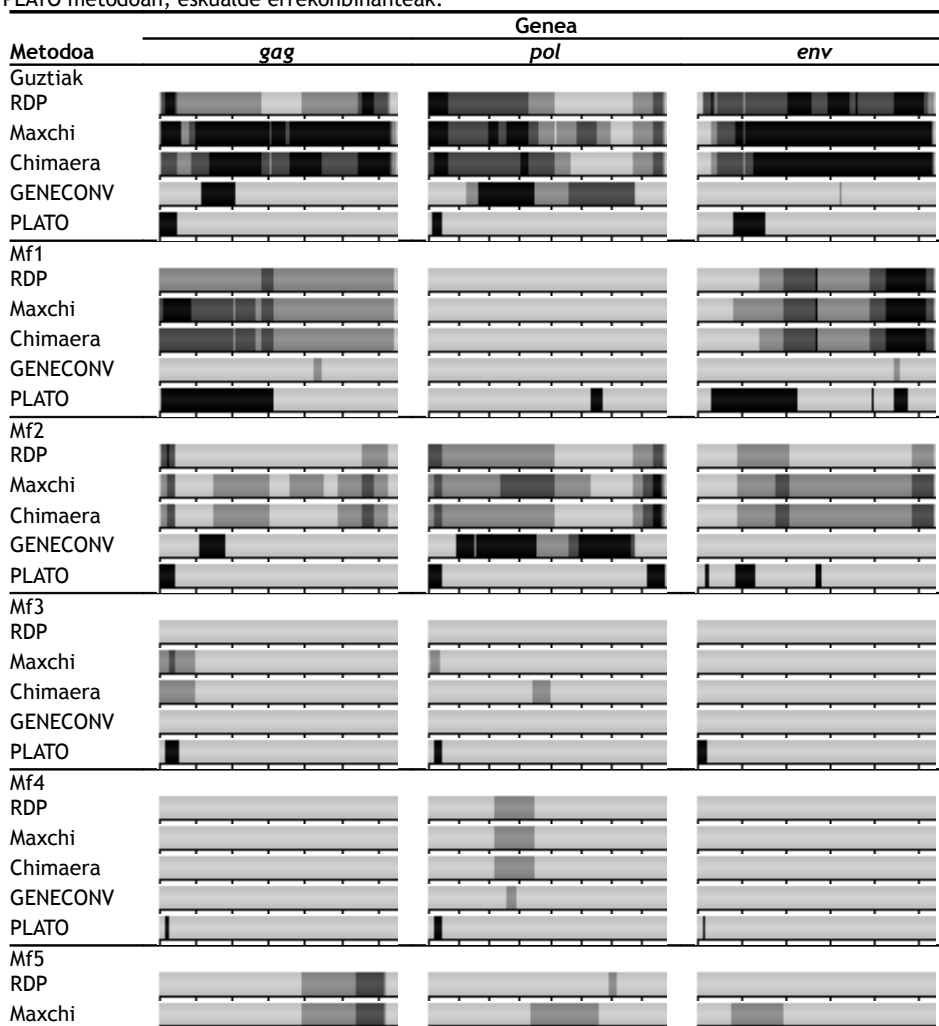
RDP, Maxchi, Chimaera eta GENECONV metodoetan beltz kolorearen intentsitateak errekonbinazio-gertaeren kopurua adierazten du. PLATO metodoan, eskualde errekonbinanteak.

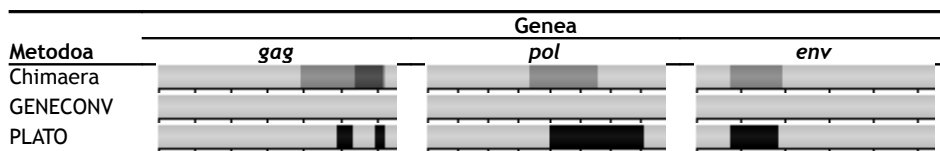




IV-6. Taula gehigarria: Aztertutako arratoiaren ERVen errekonbinazio intragenikoaren detekzioa.

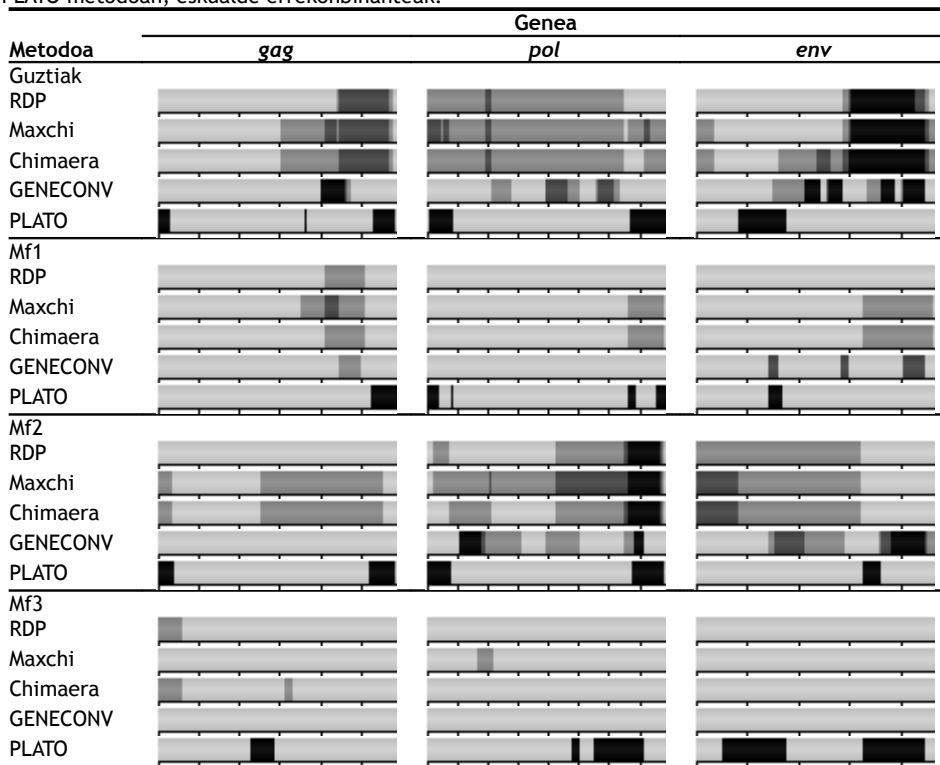
RDP, Maxchi, Chimaera eta GENECONV metodoetan beltz kolorearen intentsitateak errekonbinazio-gertaeren kopurua adierazten du. PLATO metodoan, eskualde errekonbinanteak.





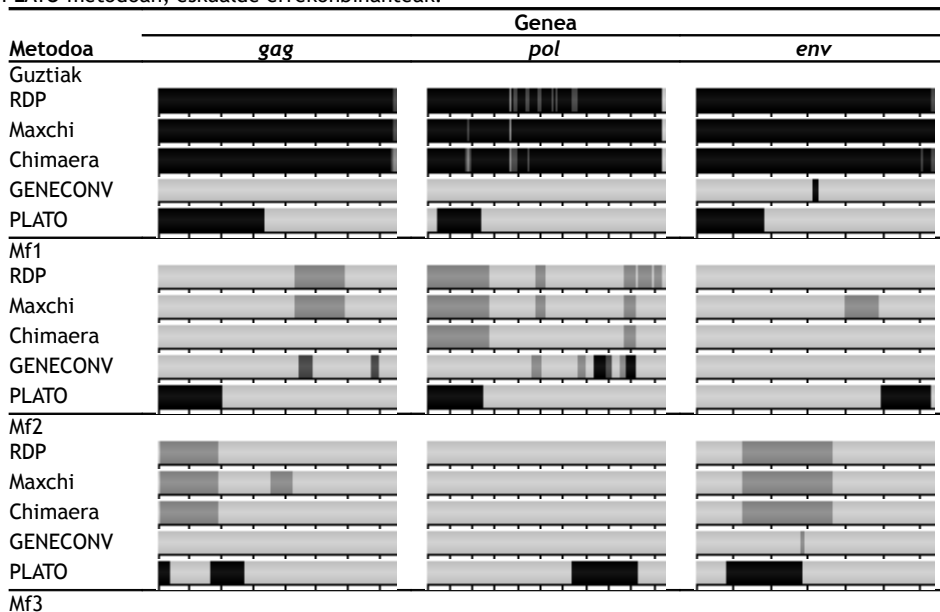
IV-7. Taula gehigarria: Aztertutako txakurraren ERVen errekonbinazio intragenikoaren detekzioa.

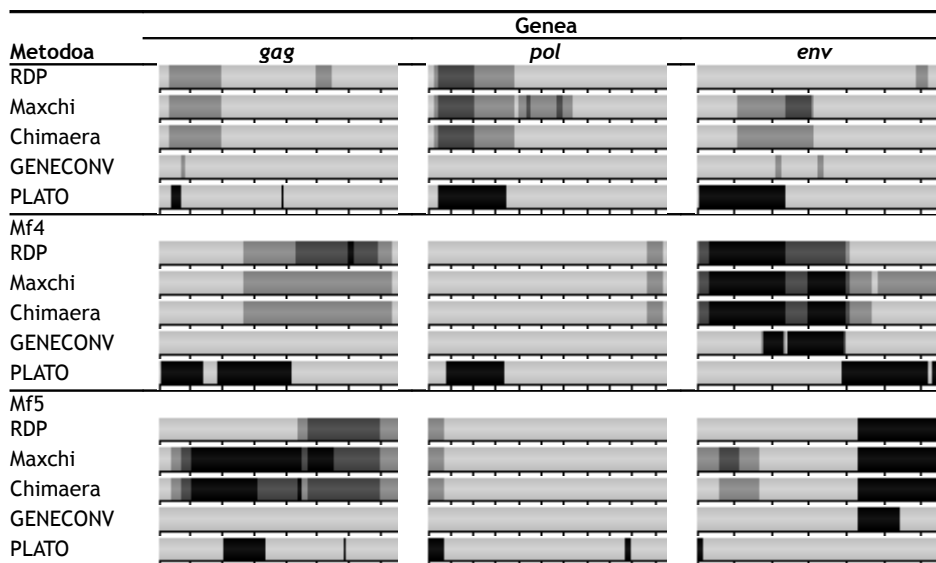
RDP, Maxchi, Chimaera eta GENECONV metodoetan beltz kolorearen intentsitateak errekonbinazio-gertaeren kopurua adierazten du. PLATO metodoan, eskualde errekonbinanteak.



IV-8. Taula gehigarria: Aztertutako behiaren ERVen errekonbinazio intragenikoaren detekzioa.

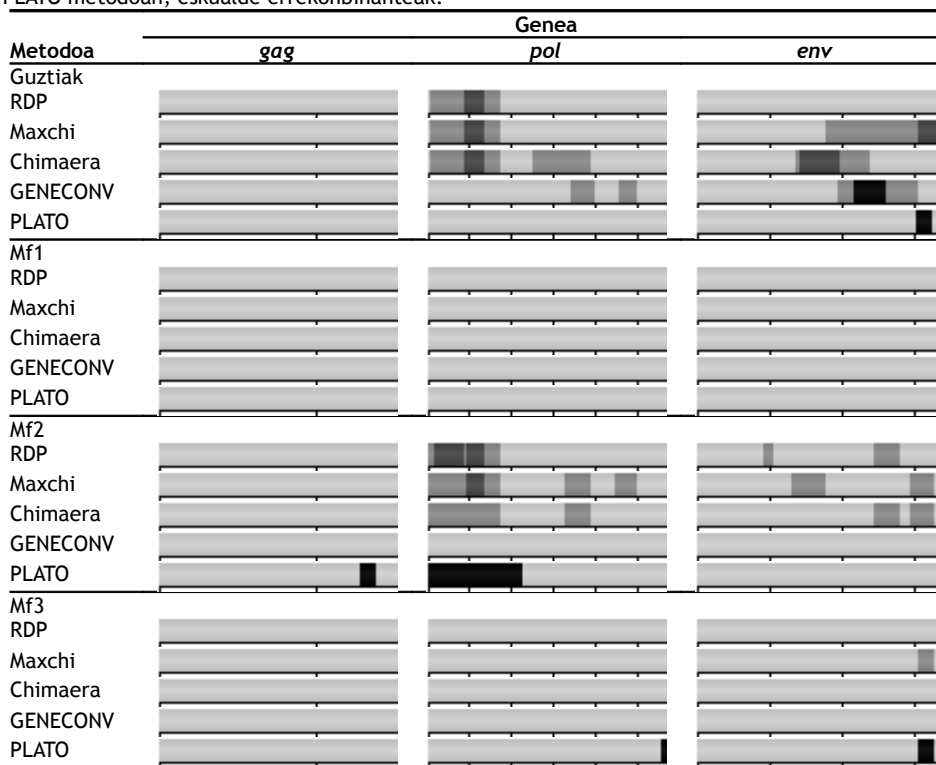
RDP, Maxchi, Chimaera eta GENECONV metodoetan beltz kolorearen intentsitateak errekonbinazio-gertaeren kopurua adierazten du. PLATO metodoan, eskualde errekonbinanteak.



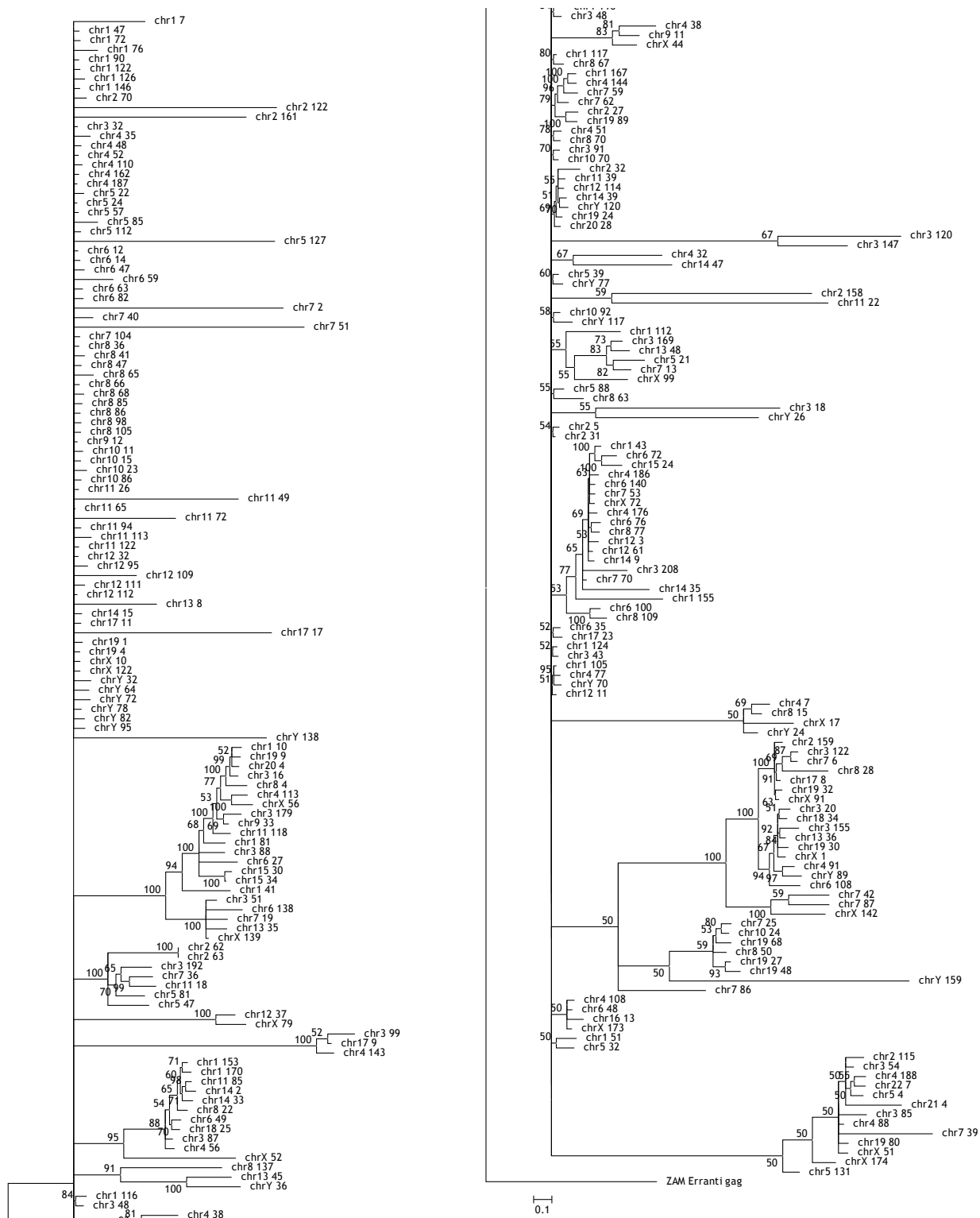


IV-9. Taula gehigarria: Aztertutako zaldiaren ERVen errekonbinazio intragenikoaren detekzioa.

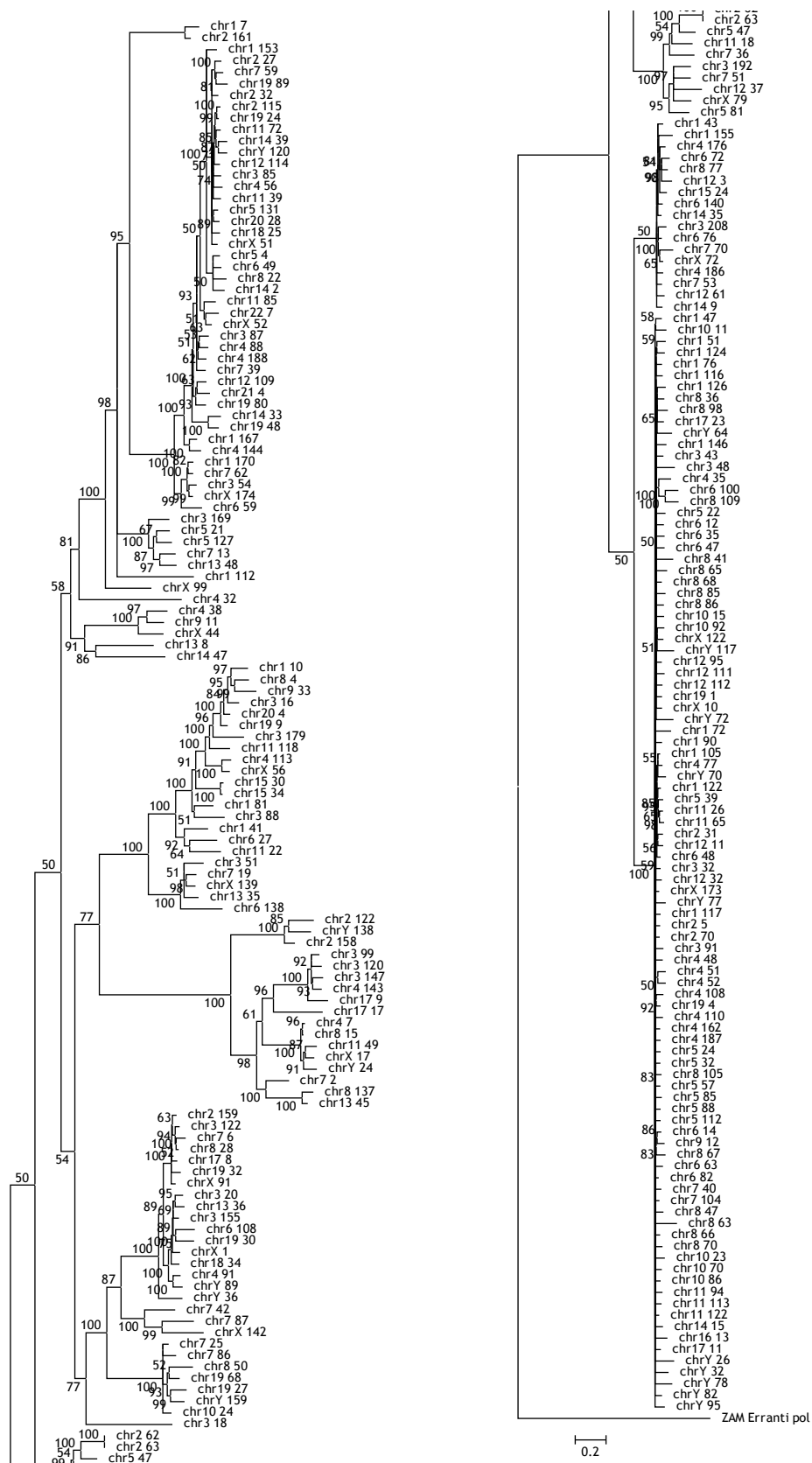
RDP, Maxchi, Chimaera eta GENECONV metodoetan beltz kolorearen intentsitateak errekonbinazio-gertaeren kopurua adierazten du. PLATO metodoan, eskualde errekonbinanteak.



Mamalioen erretrobirus endogenoen (ERV) dinamika ebolutiboa

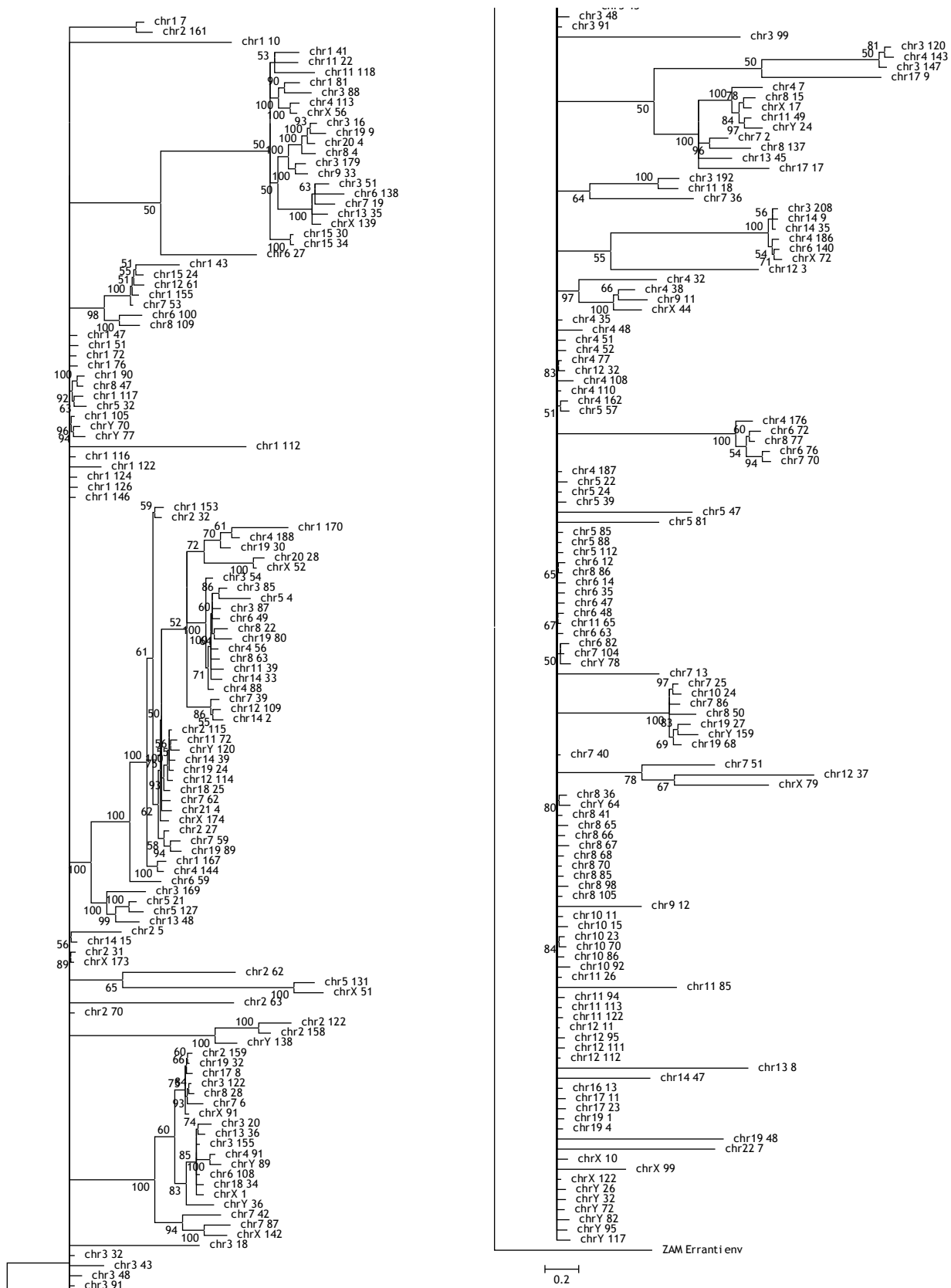


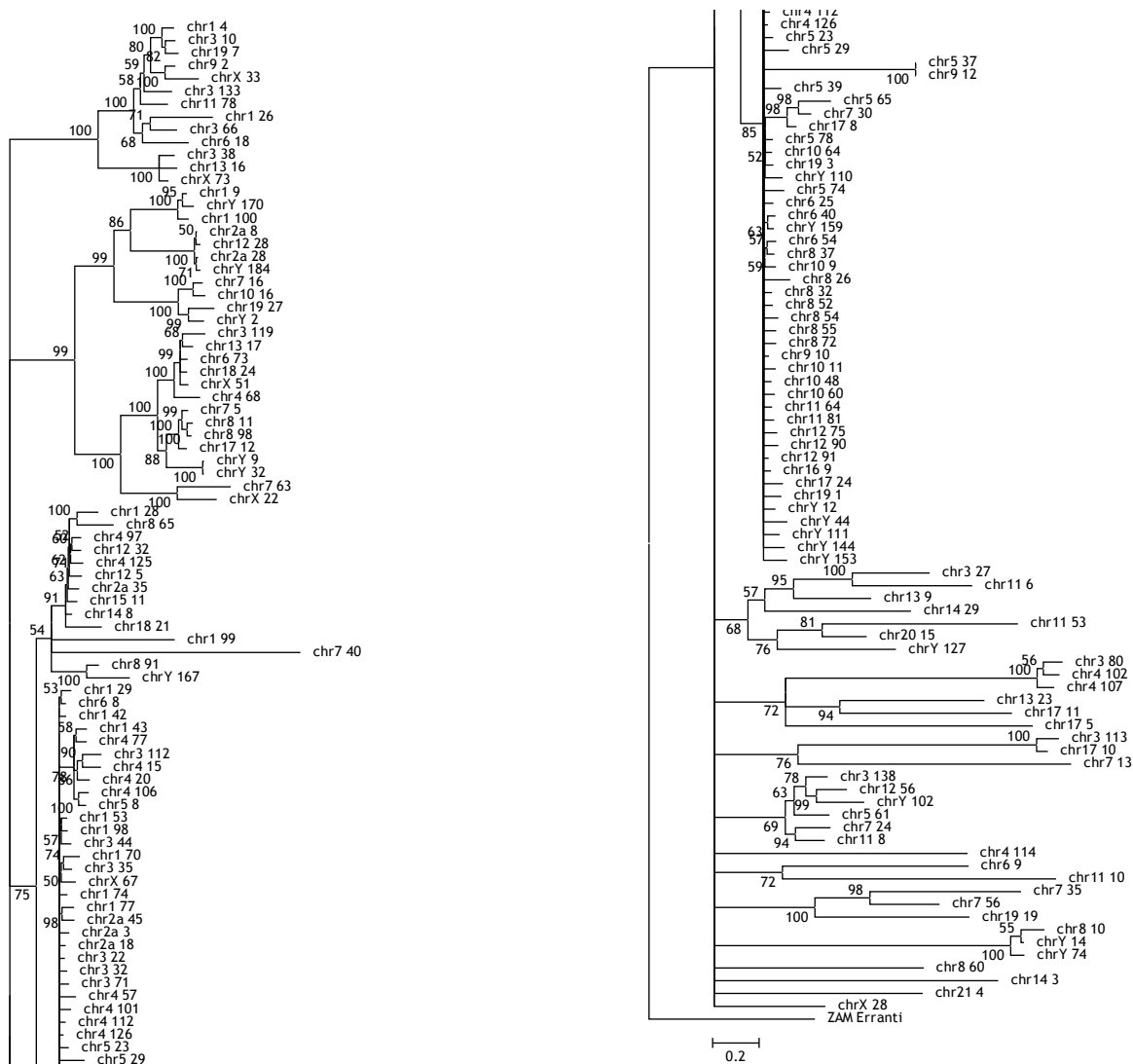
IV-1. irudi gehigarria: Aztertutako gizakiaren ERV-en gag genearen zuhaitz filogenetiko errota. Inferentzia Bayestarra, 10^6 belaunaldi, eredu ez zen finkatu.



IV-2. irudi gehigarria: Aztertutako gizakiaren ERVen *pol* genearen zuhaitz filogenetiko errotoa. Inferentzia Bayestarra, 10^6 belaunaldi, erdua ez zen finkatu.

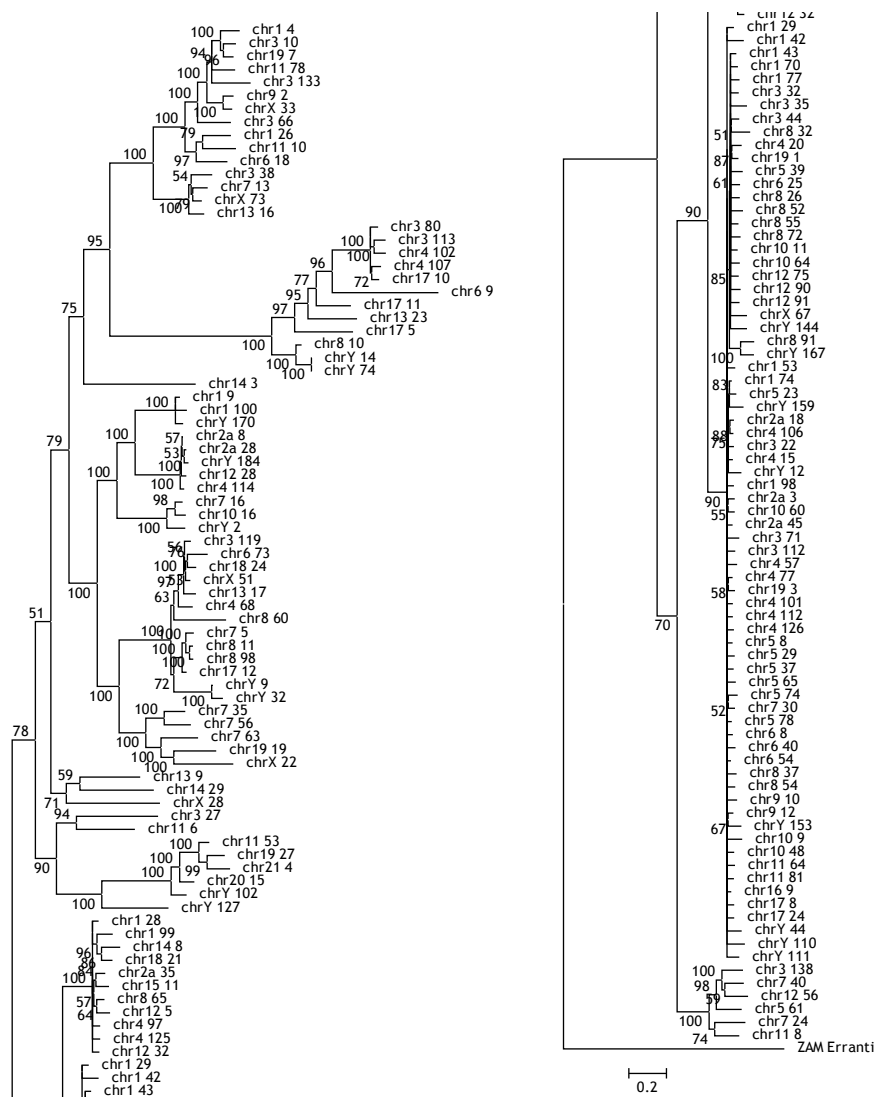
Mamaloien erretrubirus endogenoen (ERV) dinamika ebolutiboa



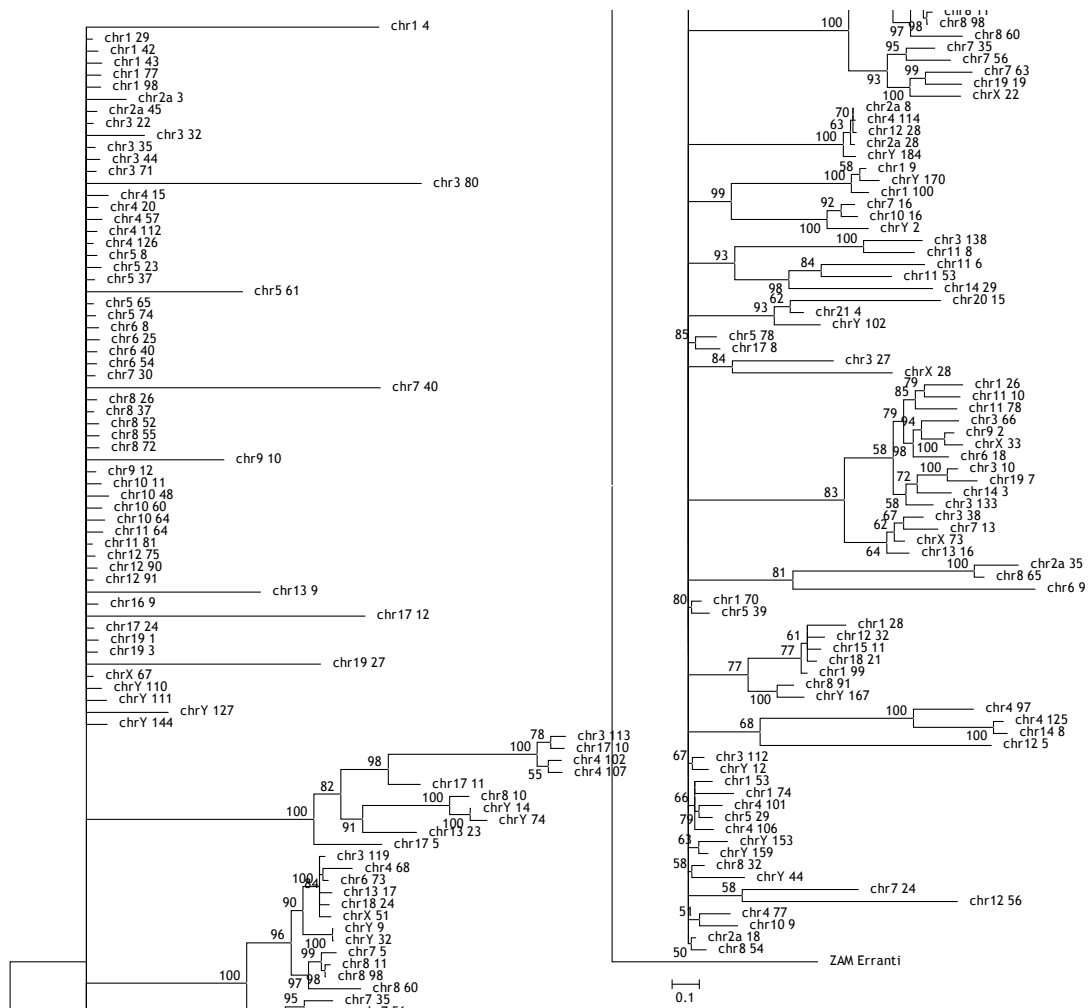


IV-4. irudi gehigarria: Aztertutako txinpantzearen ERVen gag genearen zuhaitz filogenetiko errota. Inferentzia Bayestarra, 10^6 belaunaldi, eredu ez zen finkatu.

Mamaloien erretrubirus endogenoen (ERV) dinamika ebolutiboa

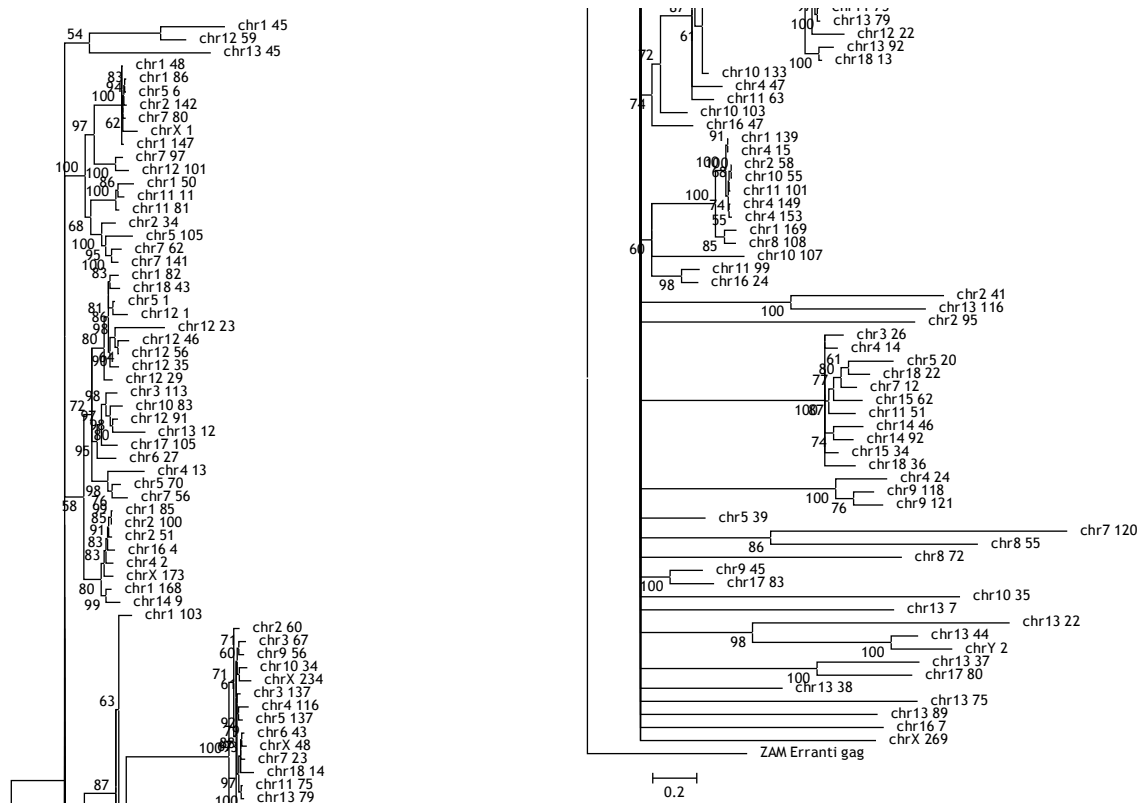


IV-5. irudi gehigarria: Aztertutako txinpantzearen ERVen *pol* genearen zuhaitz filogenetiko errotoa. Inferentzia Bayestarra, 10^6 belaunaldi, eredia ez zen finkatu.

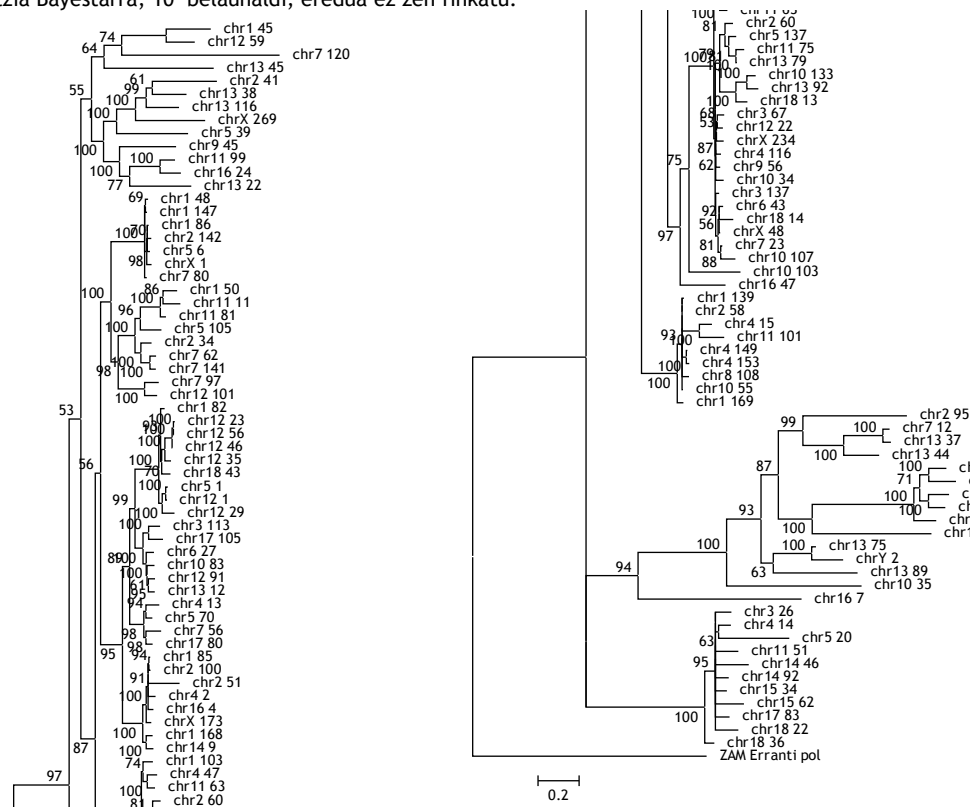


IV-6. irudi gehigarria: Aztertutako txinpantzearen ERVen *env* genearen zuhaitz filogenetiko errota. Inferentzia Bayestarra, 10^6 belaunaldi, eredu ez zen finkatu.

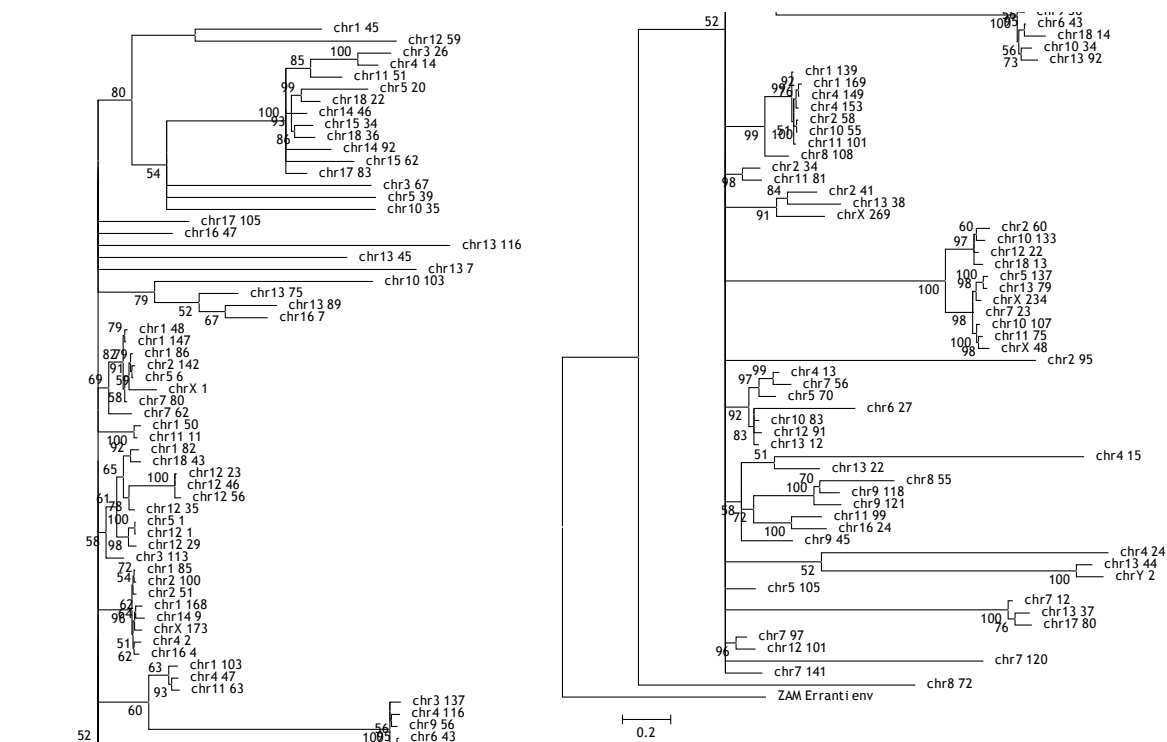
Mamaloien erretroribirus endogenoen (ERV) dinamika ebolutiboa



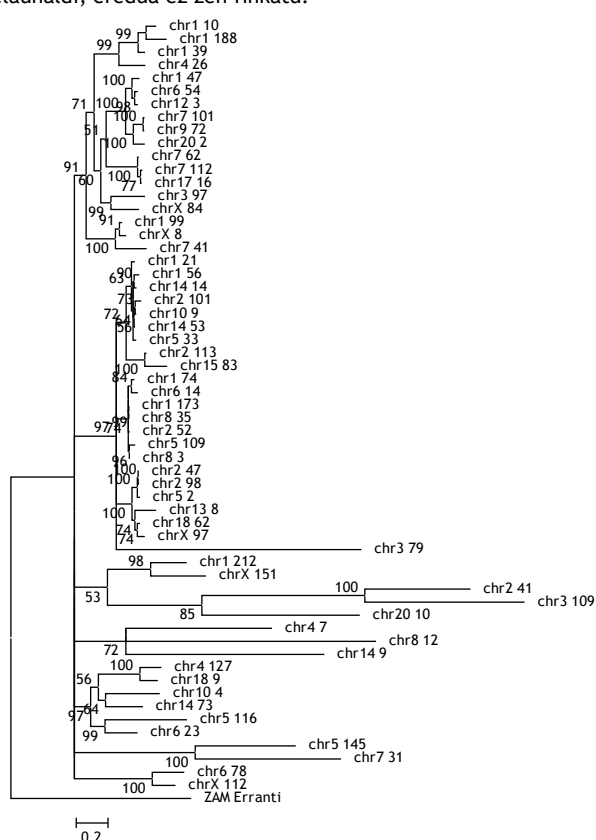
IV-7. Irudi gehigarria: Aztertutako saguaren ERV-en *gag* genearen zuhaitz filogenetiko errota. Inferentzia Bayestarra, 10^6 belaunaldi, eredu ez zen finkatu.



IV-8. irudi gehigarria: Aztertutako saguaren ERV-en *pol* genearen zuhaitz filogenetiko errota. Inferentzia Bayestarra, 10^6 belaunaldi, eredu ez zen finkatu.

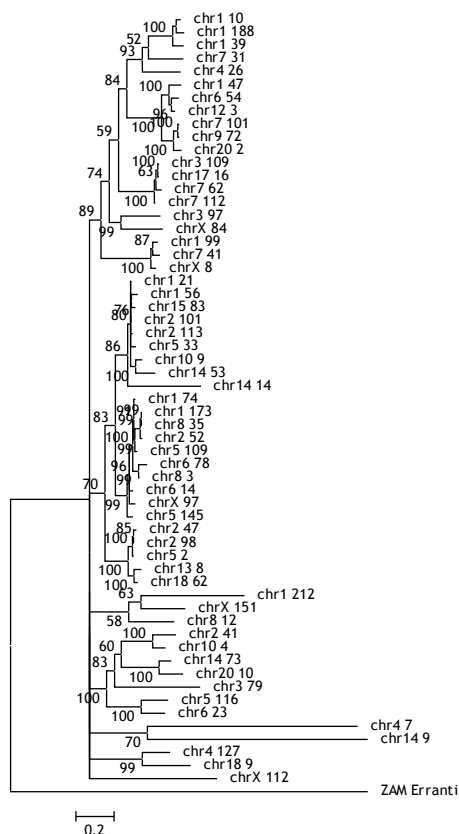


IV-9. irudi gehigarria: Aztertutako saguaren ERVen *env* genearen zuhaitz filogenetiko errotoa. Inferentzia Bayestarra, 10^6 belaunaldi, eredua ez zen finkatu.

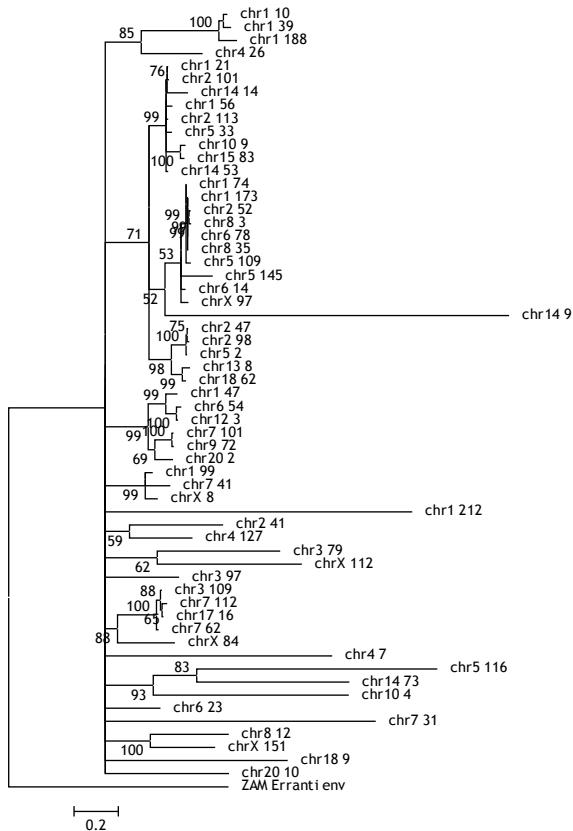


IV-10. irudi gehigarria: Aztertutako arratoiaren ERVen *gag* genearen zuhaitz filogenetiko errotoa. Inferentzia Bayestarra, 10^6 belaunaldi, eredua ez zen finkatu.

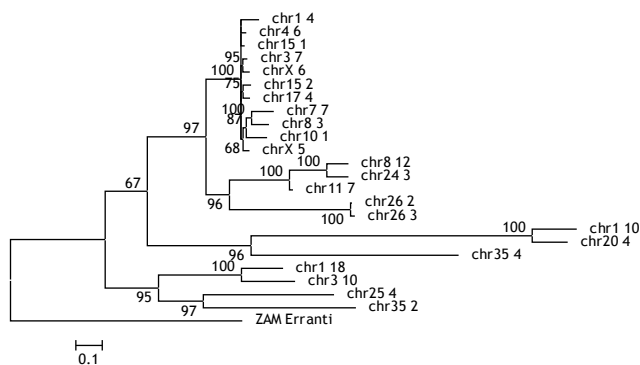
Mamaloien erretrobirus endogenoen (ERV) dinamika ebolutiboa



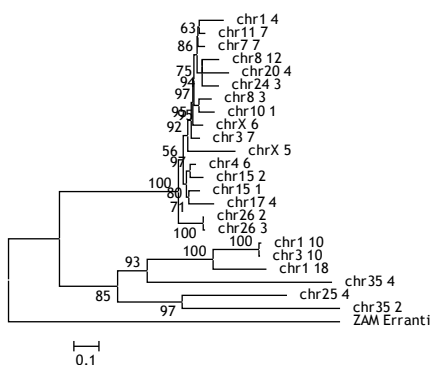
IV-11. irudi gehigarria: Aztertutako arratoiaren ERVen *pol* genearen zuhaitz filogenetiko errotua. Inferentzia Bayestarra, 10^6 belaunaldi, erdua ez zen finkatu.



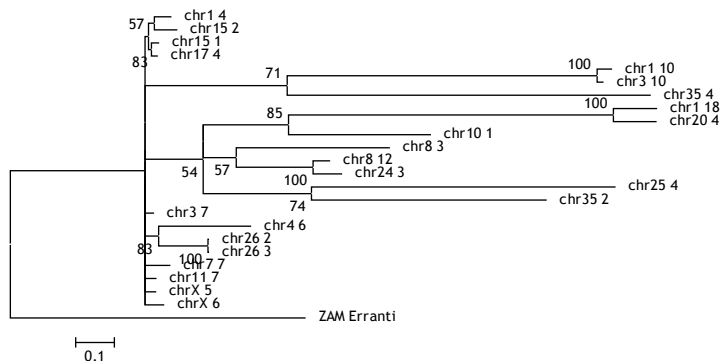
IV-12. irudi gehigarria: Aztertutako arratoiaren ERVen *env* genearen zuhaitz filogenetiko errotua. Inferentzia Bayestarra, 10^6 belaunaldi, erdua ez zen finkatu.



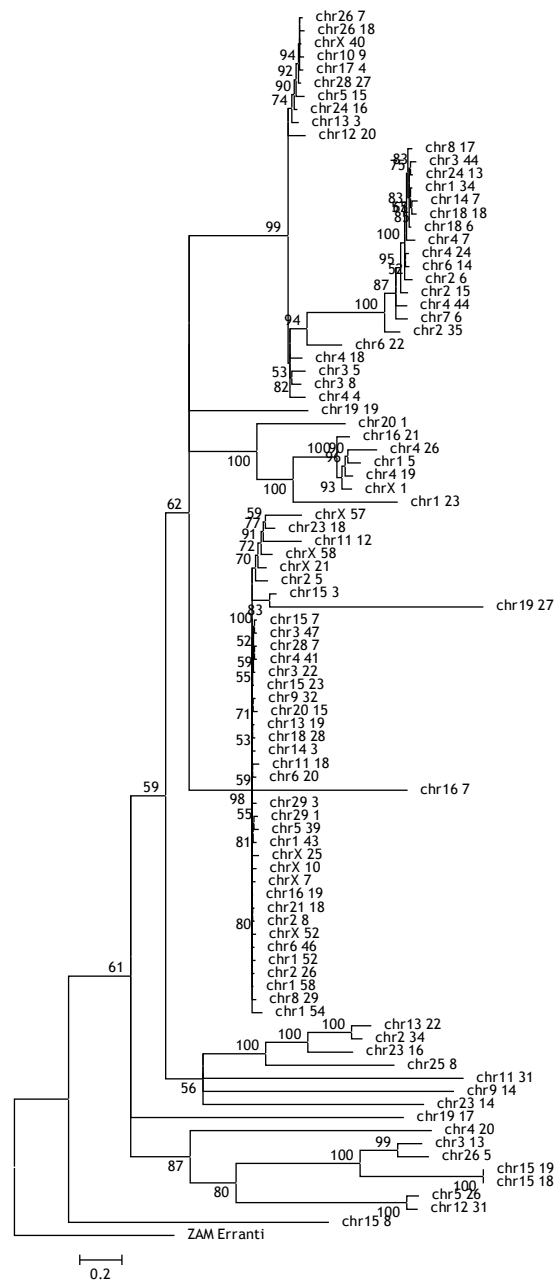
IV-13. irudi gehigarria: Aztertutako txakurraren ERVen *gag* genearen zuhaitz filogenetiko errotua. Inferentzia Bayestarra, 10^6 belaunaldi, eredia ez zen finkatu.



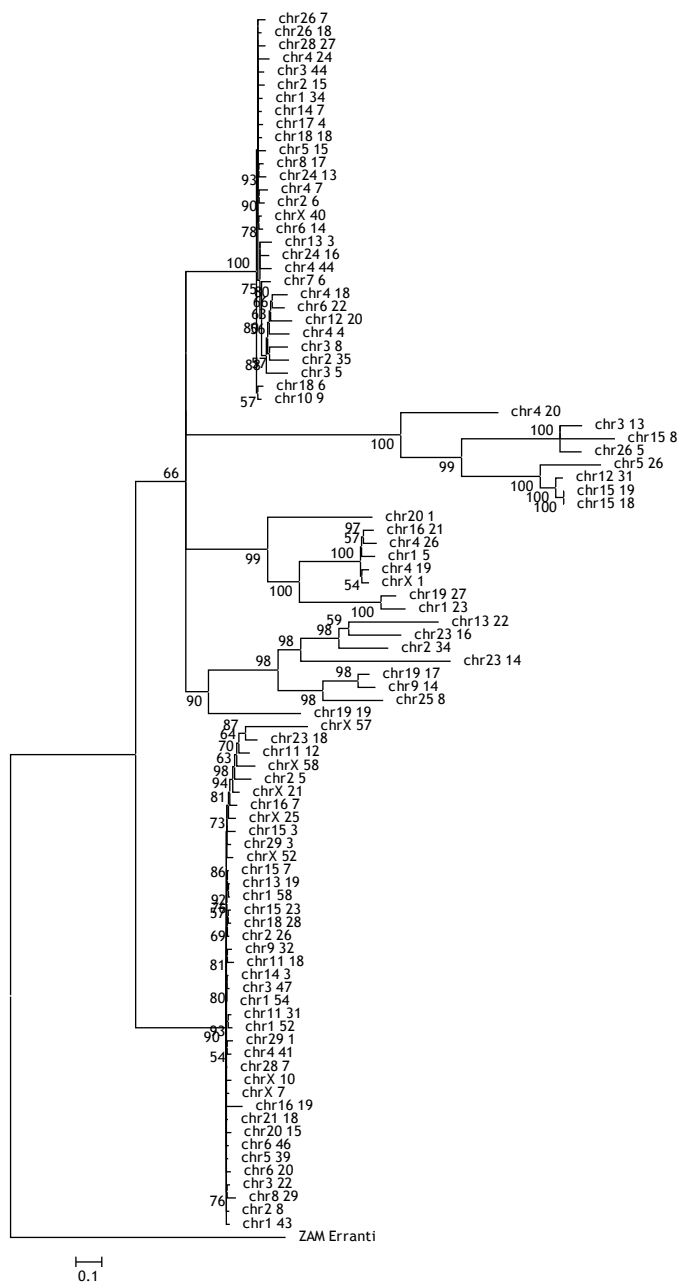
IV-14. irudi gehigarria: Aztertutako txakurraren ERVen *pol* genearen zuhaitz filogenetiko errotua. Inferentzia Bayestarra, 10^6 belaunaldi, eredia ez zen finkatu.



IV-15. irudi gehigarria: Aztertutako txakurraren ERVen *env* genearen zuhaitz filogenetiko errotua. Inferentzia Bayestarra, 10^6 belaunaldi, eredia ez zen finkatu.

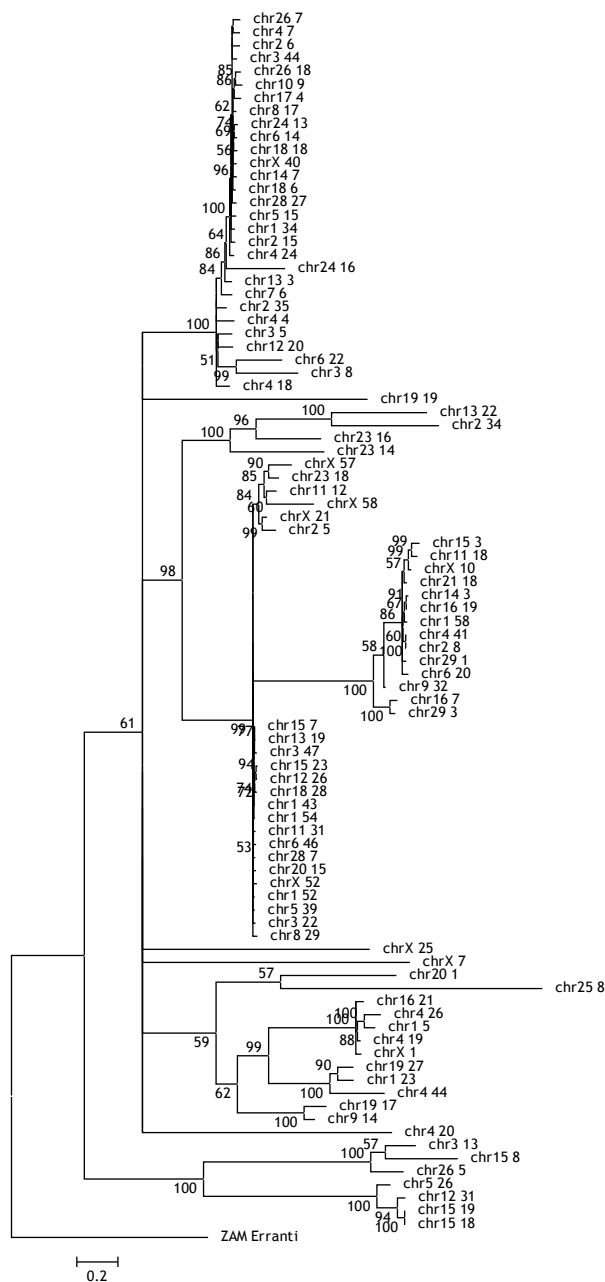


IV-16. irudi gehigarria: Aztertutako behiaren ERVen *gag* genearen zuhaitz filogenetiko errota. Inferentzia Bayestarra, 10^6 belaunaldi, eredua ez zen finkatu.

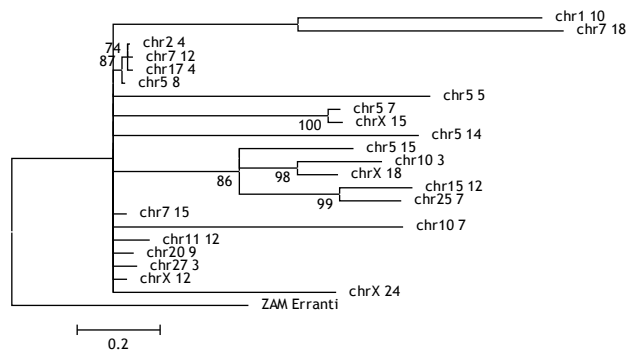


IV-17. irudi gehigarria: Aztertutako behiaren ERVen *pol* genearen zuhaitz filogenetiko errotoa. Inferentzia Bayestarra, 10^6 belaunaldi, eredu ez zen finkatu.

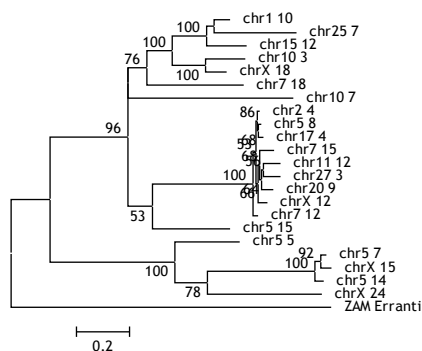
Mamalioen erretrobirus endogenoen (ERV) dinamika ebolutiboa



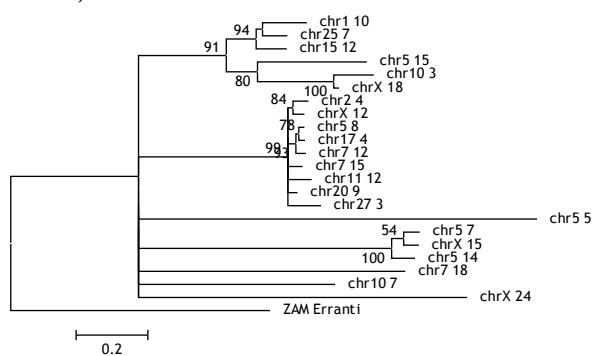
IV-18. irudi gehigarria: Aztertutako behiaren ERVen *env* genearen zuhaitz filogenetiko errotua. Inferentzia Bayestarra, 10^6 belaunaldi, eredua ez zen finkatu.



IV-19. irudi gehigarria: Aztertutako zaldiaren ERVen *gag* genearen zuhaitz filogenetiko errotua. Inferentzia Bayestarra, 10^6 belaunaldi, eredua ez zen finkatu.



IV-20. irudi gehigarria: Aztertutako zaldiaren ERVen *pol* genearen zuhaitz filogenetiko errotua. Inferentzia Bayestarra, 10^6 belaunaldi, eredia ez zen finkatu.



IV-21. irudi gehigarria: Aztertutako zaldiaren ERVen *env* genearen zuhaitz filogenetiko errotua. Inferentzia Bayestarra, 10^6 belaunaldi, eredia ez zen finkatu.

**V. Genes around bovine
endogenous retroviruses
(BoERVs) are related to host
defense and histone functions**

V.I. Introduction

Endogenous retroviruses (ERVs) are the proviral phase of exogenous retroviruses that were integrated into the host germ line and have remained in the host genome for generations (Tristem, 2000). One of the main characteristics of retroviruses is their ability to integrate into the host genome due to their replication mechanism. This step is carried out by the integrase (IN) enzyme (Vogt, 1997).

Integration was originally assumed to be a random process (Vogt, 1997). However, it was observed that in vitro IN could join retroviral DNA to naked target DNA and in vivo this process seemed to be influenced by host components (Katz et al., 1998). Moreover, some studies which have analysed the integration targeting of different exogenous retroviruses in the human genome, such as HIV, murine leukaemia virus (MLV) and avian sarcoma-leukosis virus (ASLV) have shown differently favoured target sites. For instance, HIV favours integration into transcript units, MLV prefers the 5' end of transcription units and ASLV shows a random integration pattern (Bushman et al., 2005).

In general, transposable elements (TE) or a subclass of them have been searched in coding genes on the genomes of humans (Nekrutenko & Li, 2001; Medstrand et al., 2002; van de Lagemaat et al., 2003; Jordan et al., 2003), mice (van de Lagemaat et al., 2003; DeBarry et al., 2006), *Arabidopsis* (Lockton & Gaut, 2009) and cows (Almeida et al., 2007) and their gene structure contribution has been studied. In addition, genes around LTR retrotransposons have been studied in *C. elegans* (Ganko et al., 2003) and *Drosophila* (Ganko et al., 2006).

On the whole, TEs have a minor contribution to protein coding regions: 4% in humans (Nekrutenko & Li, 2001); and the contribution of LTR retrotransposons is 1.5% in mice (DeBarry et al., 2006), 1.9% in *Drosophila* (Ganko et al., 2006) and 0.6% in *C. elegans* (Ganko et al., 2003). However, TEs have a remarkable presence in UTR regions; with TEs being present in 27% of human UTRs and 18% of mice (van de Lagemaat et al., 2003), and promoter regions 25% in humans (Jordan et al., 2003).

Moreover, in order to find functional trends in TE insertions, the genes related with TEs were assigned to Gene Ontology (GO) categories to find overrepresented functions. In humans and mice, GO categories of genes with TE contributed promoters were related with stress response, defence and response to external stimuli (van de Lagemaat et al., 2003); in mice, in genes associated with LTR retrotransposon, GO categories related with metabolic functions were overrepresented (DeBarry et al., 2006); in *Arabidopsis*, in protein coding regions with TEs, GO terms such as “kinase activity” or “transferase activity” were overrepresented (Lockton & Gaut, 2009); and in *Drosophila*, genes near LTR retrotransposons were related to GO terms such as “development” and “behaviour” (Ganko et al., 2006).

Bos taurus is an interesting species, on the one hand because it is the first ruminant to be sequenced and on the other for its economic implications (Elsik et al., 2009). In addition, several bovine ERV families have recently been detected and characterised in the cow's genome [Chapter II].

In cows, different TE derived sequences have been detected in gene introns and, in a lesser amount, in exons and promoter regions (Almeida et al., 2007). Particularly, LTR elements (e.g. ERVs) show a significant antisense insertion bias in bovine gene introns (Almeida et al., 2007). In addition, it has been proposed that this bias could be related to the survival of ERVs in gene introns (van de Lagemaat et al., 2006).

Moreover, the insertion of ERVs could shape the expression pattern of the genome. Some studies have shown the effect of ERV's LTR in the expression of genes (Jern & Coffin, 2008). These authors also suggest that the effect of ERV LTR could extend over distances of up to 100 kb

(Jern & Coffin, 2008).

Intriguing points in need of clarification are whether ERVs have integration site preferences or insertion biases, and the impact of such events on the host-genome. So, in this study we seek and characterise genes around bovine ERVs and the insertion direction with the aim of improving our knowledge about ERV integration sites and the potential functional implications for the cow genome.

V.2. Materials and methods

In this study version Btau_3.1 of the cow genome (*Bos taurus*) and the genes described in it were used (Elsik et al., 2009). The gene database for this version was downloaded from NCBI's server (ftp://ftp.ncbi.nih.gov/genomes/MapView/Bos_taurus/sequence/BUILD.3.1/initial_release/) and stored in MySQL format.

In order to characterise genes which are situated around selected [Chapter II] bovine endogenous retroviruses (BoERVs), described bovine genes and detected BoERVs were connected by home-made scripts written in PHP. This analysis was performed with 4 BoERVs datasets to compare the different sorts of BoERVs detected (Table V-1): a dataset with all detected BoERVs (Dataset I); a dataset with BoERVs with the two LTRs and the pol gene (Dataset II); a dataset with BoERVs detected by at least two of the three methods used in the detection of BoERVs (Dataset III); and a dataset with BoERVs whose gag, pol and env genes were detected with BLAST (Dataset IV). These different datasets were built up with the aim of avoiding any potential bias due to BoERVs detection, making the datasets more and more conservative. Thus, it may be possible to learn whether the results of one dataset hold when the criteria of the BoERVs selected changes.

Table V-1: Features of used datasets.

Dataset name	Number of ERVs	Features and observations
Dataset I	13622	All detected ERVs.
Dataset II	5325	Complete ERVs. ERVs which may be able to replicate.
Dataset III	1532	ERVs detected by at least two methods. ERVs which should not be false positives.
Dataset IV	93	ERVs whose 3 genes were detected. Most conserved ERVs.

In each dataset a search was performed to find genes around BoERVs in different positions and ranges: BoERVs which were included in genes or genes which were included in BoERVs and genes which were upstream or downstream of BoERVs up to 5, 50 and 100 kilobases, since the effects of ERV LTR may extend for distances of up to 100kb (Jern & Coffin, 2008). The BoERVs included in genes were classified depending on their location in introns or exons.

A χ^2 test was used to detect significant differences between the distribution of sense and antisense insertion of BoERVs as in Almeida et al (2007). The orientation of BoERV was assessed by comparing the direction of BoERV with the direction of adjacent genes and the expected values were 50% for each sense.

In order to identify overrepresented Gene Ontology terms in the genes from the datasets and ranges used the FatiGO Compare tool by Babelomics (Al-Shahrour et al., 2006) was used, because of the availability of an annotation for cows in this tool. (The options used were: gene list versus rest of genome; GO – biological process databases with default parameters; Fisher exact test). For all the analyses performed only terms with adjusted p-values < 0.05 were retrieved. The adjusted p-values were calculated by FatiGO using the FDR procedure (Al-Shahrour et al., 2006).

To characterise the genes from overrepresented Gene Ontology terms, it was determined whether they were pseudogenes or predicted genes. To characterise BoERVs related with these terms, the integrity of the structure was determined and the insertion time estimated, using the LTR divergence and a substitution rate of 2.3-5.0x10⁻⁹ substitutions/site/year (Johnson & Coffin, 1999). For these analyses home-made scripts written in PHP were used.

V.3. Results

The number of genes detected around bovine endogenous retroviruses (BoERV) changed depending on the BoERV set and range (Table V-2). In the range called “included”, i.e. BoERVs that were within genes or genes that were in BoERVs, there were quite a number of BoERV/gene associations within gene introns (2931 in Dataset I, 1109 in Dataset II, 275 in Dataset III and 24 in Dataset IV). In the case of genes upstream or downstream of BoERVs the number of associations increased with the number and distance of BoERVs. For example, in Dataset I (where 13622 ERVs were used) at “5kb upstream” there were 761 associations and at “100kb upstream” there were 12581, whereas in Dataset IV (93 ERVs used) the figures were 5 and 169 respectively.

Table V-2: Overview of gene and BoERVs associations in each range and BoERV data-set.
Upstream or downstream from BoERV.

Range			Location					
	Included		Upstream (Kb)			Downstream (Kb)		
	Gene in BoERV	BoERV in intron	5	50	100	5	50	100
Dataset I (13622 ERVs)								
Gene/BoERV associations	439	2931	761	6585	12581	734	6676	12652
Genes (15006 uniques)	425	1924	728	5438	9072	704	5504	9089
BoERVs (9565 uniques)	386	2891	714	4175	6167	686	4233	6114
Dataset II (5325 ERVs)								
Gene/BoERV associations	185	1109	298	2613	4917	296	2643	5054
Genes (8330 uniques)	180	888	293	2392	4268	288	2446	4369
BoERVs (3749 uniques)	164	1091	270	1617	2391	282	1690	2451
Dataset III (1532 ERVs)								
Gene/BoERV associations	43	275	110	951	1868	89	908	1744
Genes (2993 uniques)	42	236	100	797	1500	84	773	1429
BoERVs (1081 uniques)	37	263	101	573	797	88	552	769
Dataset IV (93 ERVs)								
Gene/BoERV associations	4	24	5	87	169	8	74	137
Genes (328 uniques)	4	24	5	87	169	8	74	137
BoERVs (79 uniques)	4	23	5	48	62	7	45	64

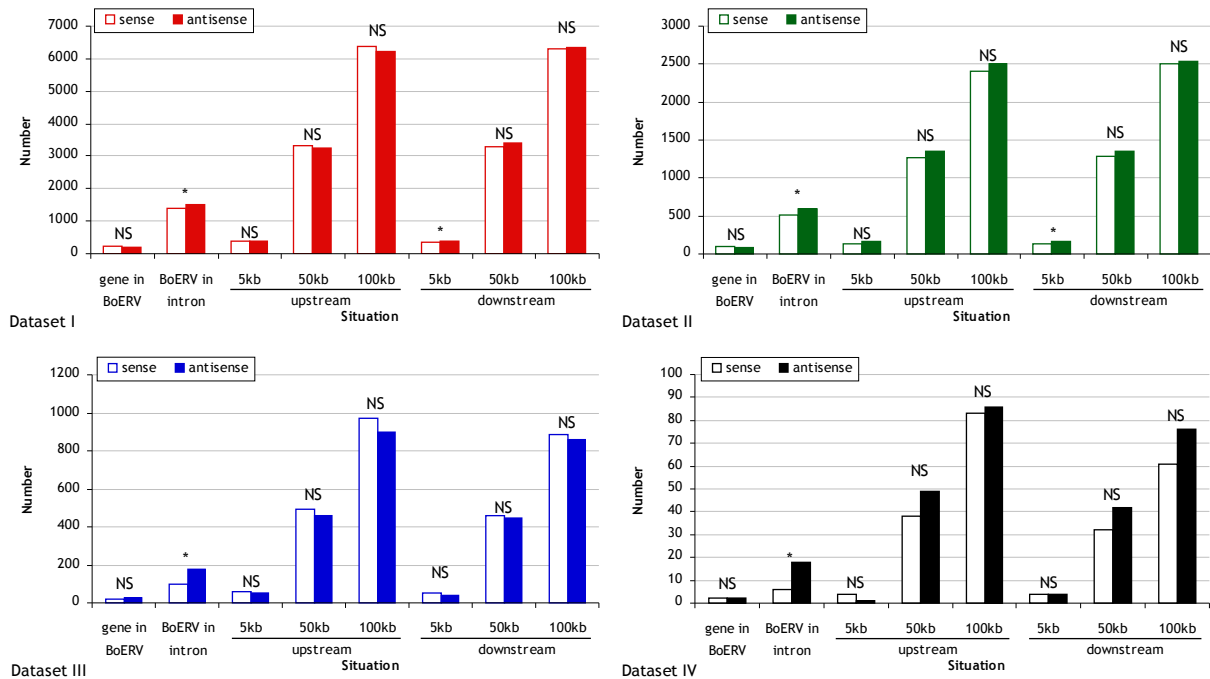


Figure V-1: Insertion orientation of BoERVs in each range and data-set.
NS, no significant; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

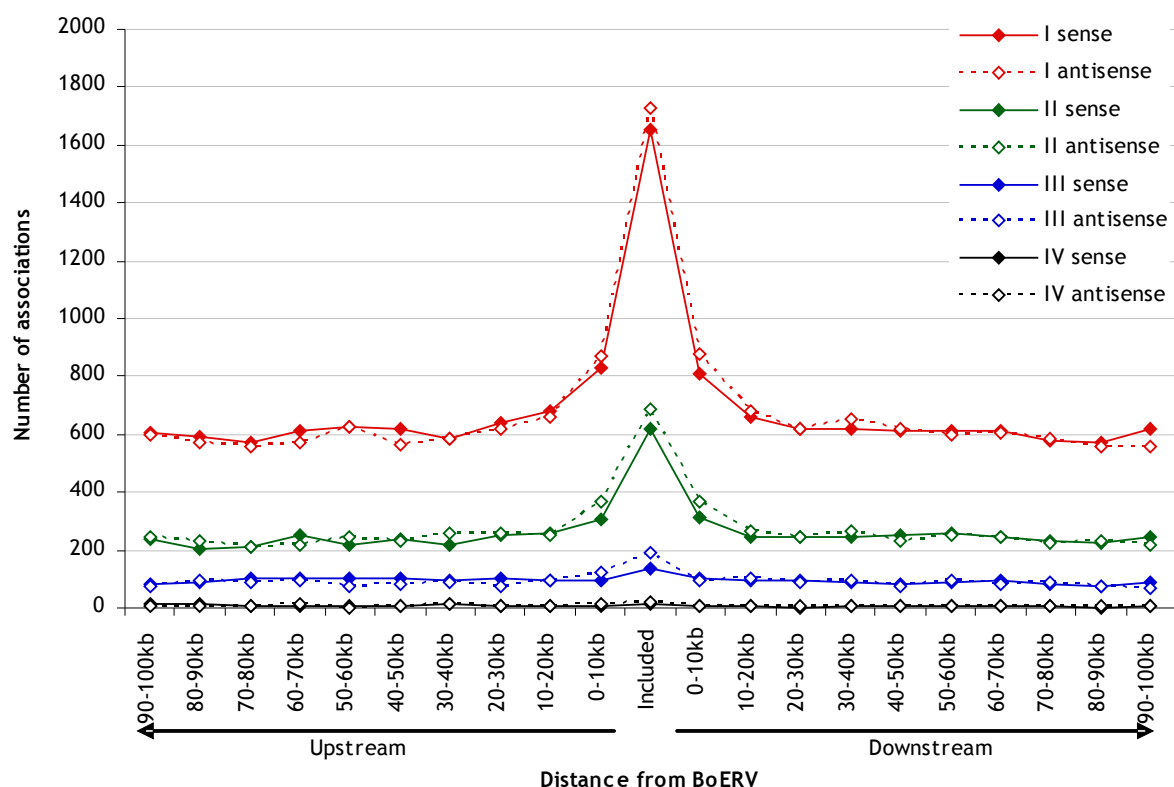


Figure V-2: Number of associations of BoERVs and genes in each dataset. Included, BoERVs within genes or genes in BoERVs. I: Dataset I; II: Dataset II; III: Dataset III; IV: Dataset IV.

Regarding the insertion bias (Figure V-1), all the datasets used had similar results. In the included range, the BoERVs inserted in gene introns were inserted in an antisense orientation significantly in all datasets and in the case of genes within BoERVs there was no insertion bias. Overall, the BoERVs that were near genes did not show any sense or antisense insertion bias. However, in the case of genes “5kb downstream” of BoERVs, in BoERVs from Dataset I and Dataset II they were inserted in an antisense orientation significantly.

The number of BoERV/gene associations in the sense and antisense directions differed over the distance from the BoERVs (Figure V-2). In all the datasets analysed, the highest number of associations was found in the included ranges, that is, in genes within BoERVs or BoERVs within genes. Most associations happened in the range between 0 and 10 kb, upstream or downstream, starting from the BoERV. From there, the number of associations was stable. There was no clear trend of insertion direction except in the “included” location and 0-10kb ranges, where antisense associations were more frequent.

On the whole, around 20% of the genes near BoERVs were pseudogenes and around half were predicted genes (Table V-3), although there were differences in each dataset or range. The highest values of pseudogenes and predicted genes were found in genes within BoERVs and the lowest, in general, in BoERVs within introns. The overall figure for the annotated gene descriptions of the versions used was 13.88% for pseudogenes and 40.40% for predicted genes.

The genes within and around BoERVs were assigned to a biological process according to Gene Ontology. The overrepresented categories whose adjusted p-values were <0.05 (Supplementary tables V-1 to V-4) were taken (Table V-4). There were overrepresented terms in some of the ranges studied in each dataset: all ranges in Dataset I except “BoERVs in introns”; in Dataset II, there were such terms in “genes within BoERVs”, “50kb upstream” and “50kb and 100kb downstream”; in Dataset III, they were found in all ranges except “BoERVs in intron” and 5kb in both directions;

and in Dataset IV in “50kb and 100kb upstream”. In general, the overrepresented categories which were repeated most in the different BoERV datasets and ranges were those related to host defence response, especially “response to virus”, and terms related with “chromosome organisation”, especially “chromatin assembly” (Table V-4).

Table V-3: Features of all genes near BoERVs, in each range and BoERV data-set.

Range	Included		Location					
			Upstream (Kb)			Downstream (Kb)		
	Gene in BoERV	BoERV in intron	5	50	100	5	50	100
Dataset I								
Genes	23.23%	67.92%	57.49%	55.18%	56.79%	52.42%	55.6%	56.93%
Predicted genes	46.18%	19.18%	23.28%	28.95%	28.21%	31.67%	28.49%	28.37%
Pseudogenes	1.18%	1.82%	0.82%	0.77%	0.83%	0.57%	0.58%	0.85%
Predicted pseudogenes	29.41%	11.02%	18.41%	15.10%	14.17%	15.34%	15.33%	13.85%
Dataset II								
Genes	22.33%	66.55%	46.42%	55.6%	56.61%	48.96%	54.83%	55.83%
Predicted genes	48.33%	19.71%	35.15%	28.18%	28%	32.29%	28.12%	28.15%
Pseudogenes	2.22%	1.58%	0.68%	0.75%	0.7%	1.04%	0.65%	0.82%
Predicted pseudogenes	27.22%	12.16%	17.75%	15.47%	14.69%	15.63%	16.4%	15.2%
Dataset III								
Genes	14.29%	39.53%	42%	56.81%	55.53%	51.19%	56.14%	54.17%
Predicted genes	59.52%	38.86%	34%	32%	29.53%	28.57%	27.04%	28.13%
Pseudogenes	4.76%	2.12%	1%	0.75%	0.01%	0%	0.65%	0.56%
Predicted pseudogenes	21.43%	19.49%	23%	16.44%	14.93%	20.24%	16.17%	17.14%
Dataset IV								
	25%	58.38%	60%	54.03%	58.59%	62.5%	55.41%	53.28%
Predicted genes	50%	16.67%	0%	27.58%	21.30%	12.5%	31.08%	31.39%
Pseudogenes	0%	0%	0%	1.15%	1.77%	0%	0%	0%
Predicted pseudogenes	25%	25%	40%	17.24%	18.34%	25%	13.51%	15.33%

Furthermore, each dataset showed certain particularities. In Dataset I (Table V-4) the “response to virus” term and terms related to it were overrepresented in all the ranges studied except “BoERVs within introns”. The number of genes involved in the “response to virus” term was different in each range, from 6 genes (“5kb downstream”) to 31 (“100kb downstream”). However, the “chromatin assembly” term was only overrepresented in ranges up to “50kb upstream” (36 and 48 related genes respectively) and 50kb (49 and 56 genes) and 100kb (66 and 83 genes) downstream.

In the case of dataset II (Table V-4), the “response to virus” term was overrepresented in “genes within BoERVs” (5 genes), 50kb in both directions (12 genes respectively) and “100kb downstream” (22 genes). The genes situated up to “50kb and 100kb downstream” from BoERVs were related to the overrepresented “chromatin assembly” (26 genes respectively) term.

In the Dataset III (Table V-4) the “response to virus” term was overrepresented in “genes within BoERVs” (4 genes), 50kb in both directions (7 genes respectively) and “50kb upstream” (6 genes). However, terms related to “chromatin assembly” were only overrepresented in “100kb upstream” (17 genes).

In Dataset IV (Table V-4), the overrepresented terms related to “chromatin assembly” were detected in the 50kb (7 genes) and 100kb (15 genes) upstream ranges.

Besides these terms or terms related to them, other terms such as “behaviour” (Dataset III, in “genes within BoERVs”) and “mitosis” (Dataset III, in “100kb downstream”) (Table V-4) emerged in this analysis.

Genes around bovine endogenous retroviruses (BoERVs) are related to host defense and histone functions

Table V-4: Overrepresented Gene Ontology terms in each range and BoERV data-set.

N, genes involved in overrepresented term. Sig, significance of overrepresented term. † Term related with term above.

	Gene in BoERV		Included BoERV in intron		Location						Downstream (kb)					
					Upstream (kb)						5			50		
	N ^a	Sig	N ^a	Sig	N ^a	Sig	N ^a	Sig	N ^a	Sig	N ^a	Sig	N ^a	Sig	N ^a	Sig
Dataset I																
response to biotic stimulus	12	***	-		12	**	48	**	-	-	43	*				
†response to other organism	11	***	-		10	*	38	**	-	-	32	**	49	*		
†response to virus	9	***	-		7	**	26	***	28	**	6	*	24	***	31	***
defense response	12	*	-		-		50	*	-	-	-	-	-	-	-	-
biosynthetic process	-	-	-		-	-	-	266	*	-	-	-	-	-	-	-
†cellular biosynthetic process	-	-	-		-	-	-	244	*	-	-	-	-	-	-	-
†macromolecule biosynthetic process	-	-	-		-	105	*	185	**	-	102	*	161	*		
†translation	14	***	-		-	84	*	152	***	-	78	*	127	*		
†DNA metabolic process	-	-	-		-	-	-	-	-	-	70	*				
cellular component	-	-	-		-	-	-	-	-	-	-	-	-	-	-	-
cellular component assembly	-	-	-		-	54	*	-	-	-	61	***	87	***		
macromolecule complex assembly	-	-	-		-	54	**	74	*	-	60	***	85	***		
†protein-DNA complex assembly	-	-	-		-	39	**	53	*	-	50	***	69	***		
DNA packaging	-	-	-		-	41		-	-	-	53	***	72	***		
†establishment and/or maintenance of chromatin architecture	-	-	-		-	-	-	-	-	-	-	-	72	***		
†chromatin assembly or disassembly	-	-	-		-	36	*	-	-	-	50	***	67	***		
†chromatin assembly	-	-	-		-	36	*	-	-	-	49	***	66	***		
organelle organization and biogenesis	-	-	-		-	-	-	-	-	-	82	**				
†chromosome organization and biogenesis	-	-	-		-	-	-	-	-	-	59	***	81	***		
†chromosome organization and biogenesis (sensu Eukaryota)	-	-	-		-	43	*	-	-	-	58	***	80	***		
cellular developmental process	-	-	-		-	-	-	-	-	-	106	*	173	*		
†cell differentiation	-	-	-		-	-	-	-	-	-	106	*	173	*		
†cell development	-	-	-		-	-	-	-	-	-	89		142	*		
Dataset II																
response to biotic stimulus	6	**	-		-	25	*	-	-	-	-	-	-	-	-	-
†response to other organism	6	**	-		-	22	**	-	-	-	-	-	-	-	-	-
†response to virus	5	***	-		-	12	*	-	-	-	13	**	22	***		
defense response	6	**	-		-	-	-	-	-	-	-	-	-	-	-	-
cellular component assembly	-	-	-		-	-	-	-	-	-	35	***	58	***		
macromolecule complex assembly	-	-	-		-	-	-	-	-	-	34	***	56	***		
†protein-DNA complex assembly	4	*	-		-	-	-	-	-	-	27	***	44	***		
DNA packaging	-	-	-		-	-	-	-	-	-	29	***	45	***		
†establishment and/or maintenance of chromatin architecture	-	-	-		-	-	-	-	-	-	29	***	45	***		
†chromatin assembly or disassembly	4	*	-		-	-	-	-	-	-	27	***	43	***		
†chromatin assembly	-	-	-		-	-	-	-	-	-	26	***	42	***		
organelle organization and biogenesis	-	-	-		-	-	-	-	-	-	40	*	63	*		
†chromosome organization and biogenesis	-	-	-		-	-	-	-	-	-	33	***	52	***		
†chromosome organization and biogenesis (sensu Eukaryota)	-	-	-		-	-	-	-	-	-	33	***	52	***		
cellular developmental process	-	-	-		-	-	-	-	-	-	-	-	98	**		
†cell differentiation	-	-	-		-	-	-	-	-	-	-	-	98	**		
†cell development	-	-	-		-	-	-	-	-	-	-	-	83	**		
Dataset III																
response to biotic stimulus	5	***	-		-	12	**	16	*	-	-	-	17	*		
†response to other organism	4	***	-		-	10	*	-	-	-	-	-	13	*		
†response to virus	4	***	-		-	7	**	9	*	-	6	*	-	-	-	-
defense response	4	**	-		-	14	**	22	**	-	-	-	18	*		
behavior (GO:0007610)	2	*	-		-	-	-	-	-	-	-	-	-	-	-	-
chromosome segregation (GO:0007059)	-	-	-		-	-	-	-	-	-	-	-	5	*		
mitotic cell cycle (GO:0000278)	-	-	-		-	-	-	-	-	-	-	-	11	*		
†M phase of mitotic cell cycle (GO:0000087)	-	-	-		-	-	-	-	-	-	-	-	11	*		
†mitosis (GO:0007067)	-	-	-		-	-	-	-	-	-	-	-	11	*		
macromolecule complex assembly	-	-	-		-	-	-	19	*	-	-	-	-	-	-	-
†protein-DNA complex assembly	-	-	-		-	-	-	17	**	-	-	-	-	-	-	-
DNA packaging	-	-	-		-	-	-	17	*	-	-	-	-	-	-	-
†establishment and/or maintenance of chromatin architecture	-	-	-		-	-	-	17	*	-	-	-	-	-	-	-
†chromatin assembly or disassembly	-	-	-		-	-	-	17	**	-	-	-	-	-	-	-
†chromatin assembly	-	-	-		-	-	-	17	**	-	-	-	-	-	-	-
chromosome organization and biogenesis	-	-	-		-	-	-	18	*	-	-	-	-	-	-	-
chromosome organization and biogenesis (sensu Eukaryota)	-	-	-		-	-	-	18	*	-	-	-	-	-	-	-

	Gene in BoERV		Included BoERV in intron		Location						Downstream (kb)					
					Upstream (kb)											
	N ^a	Sig	N ^a	Sig	5	50	100	5	50	100	5	50	100	5	50	100
Dataset IV																
biopolymer metabolic process (GO:0043283)	-	-	-	-	-	-	21	*	-	-	-	-	-	-	-	-
nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	-	-	-	-	-	-	19	**	-	-	-	-	-	-	-	-
↓DNA metabolic process	-	-	-	-	-	7	**	16	***	-	-	-	-	-	-	-
cellular component	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
cellular component assembly	-	-	-	-	-	7	***	15	***	-	-	-	-	-	-	-
macromolecule complex assembly	-	-	-	-	-	7	**	15	***	-	-	-	-	-	-	-
↓protein-DNA complex assembly	-	-	-	-	-	7	***	15	***	-	-	-	-	-	-	-
cellular component organization and biogenesis (GO:0016043)	-	-	-	-	-	-	-	16	***	-	-	-	-	-	-	-
↓DNA packaging	-	-	-	-	-	7	***	15	***	-	-	-	-	-	-	-
Establishment and/or maintenance of chromatin architecture	-	-	-	-	-	7	***	15	***	-	-	-	-	-	-	-
↓chromatin assembly or disassembly	-	-	-	-	-	7	***	15	***	-	-	-	-	-	-	-
↓chromatin assembly	-	-	-	-	-	7	***	15	***	-	-	-	-	-	-	-
organelle organization and biogenesis	-	-	-	-	-	7	**	15	***	-	-	-	-	-	-	-
↓chromosome organization and biogenesis	-	-	-	-	-	7	***	15	***	-	-	-	-	-	-	-
↓chromosome organization and biogenesis (sensu Eukaryota)	-	-	-	-	-	7	***	15	***	-	-	-	-	-	-	-
cellular developmental process	-	-	-	-	-	8	**	32	***	-	-	-	-	-	-	-
↓cell differentiation	-	-	-	-	-	8	**	17	***	-	-	-	-	-	-	-
↓cell development	-	-	-	-	-	8	**	17	***	-	-	-	-	-	-	-

* p<0,05; ** p<0,01; *** p< 0,001

^a a gene could appear in more than one GO term.

Moreover, some genes appeared in two or more datasets in the same GO term and range (Supplementary Table V-5). In the “response to virus” term, the IFN-tau-c1 gene (interferon tau c1) appeared in three datasets at “50kb upstream” and LOC618801 (similar to interferon tau) and LOC781424 (similar to interferon alpha F) at “50kb downstream”. In the “chromatin assembly” term, 8 genes, which were similar to histone cluster related genes, appeared in three datasets at “50kb upstream” and at “100kb upstream” 13 genes which were similar to histone cluster related genes and the H2AFZ gene (H2A histone family, member Z) appeared in three datasets.

Table V-5: Features of BoERVs around genes that are in overrepresented GO terms, in each range and BoERV data-set.

-, there was not overrepresented term.

Range (kb)	Included BoERV in intron / UTR		Location					
			Upstream (Kb)			Downstream (Kb)		
	Gene in BoERV		5	50	100	5	50	100
Dataset I								
BoERV number	26	-	21	205	401	6	279	395
Complete BoERVs (%)	65.38	-	71.43	51.22	53.37	66.67	50.9	54.18
Incomplete BoERVs (%)	34.62	-	28.57	48.78	46.63	33.33	49.1	45.82
% of BoERVs around genes	6.74	-	2.94	4.91	6.5	0.87	6.59	6.46
Insertion-time (MYA)	136-11	-	127-33	190-0	190-9	127-3.5	187-0	190-4
Dataset II								
BoERV number	9	-	-	15	-	-	42	100
Complete BoERVs (%)	77.78	-	-	86.67	-	-	76.19	75
Incomplete BoERVs (%)	22.22	-	-	13.33	-	-	23.81	25
% of BoERVs around genes	5.49	-	-	0.93	-	-	2.49	4.08
Insertion-time (MYA)	127-3	-	-	127-9	-	-	187-3	187-3
Dataset III								
BoERV number	6	-	-	13	30	-	2	23
Complete BoERVs (%)	66.67	-	-	69.23	73.33	-	100	43.48
Incomplete BoERVs (%)	33.33	-	-	30.77	26.67	-	0	56.52
% of BoERVs around genes	16.22	-	-	2.27	3.76	-	0.36	2.99
Insertion-time (MYA)	110-0	-	-	190-0	190-15	-	58-20	141-14
Dataset IV								
BoERV number	-	-	-	2	15	-	-	-
% of BoERVs around genes	-	-	-	2.15	16.13	-	-	-
Insertion-time (MYA)	-	-	-	41-16	41-9	-	-	-

BoERVs and genes related to overrepresented GO terms

The BoERVs related to genes from overrepresented GO terms were a minor portion of the total BoERVs used in a dataset or BoERVs which were within or around genes (Table V-5). These BoERVs showed a different degree of conservation in each dataset and range and, on the whole, their insertion times had a wide range, from recently inserted BoERVs to ancient BoERVs.

In the case of Dataset I, the structure of the BoERVs that were near genes related to overrepresented GO terms was more complete in all ranges (Table V-5). The most ancient BoERVs, inserted 190 MYA, were found in the 50kb and 100kb ranges and the newest, most recently inserted, in the 50kb downstream range. In Dataset II BoERV structure was conserved in all ranges. The integration of the oldest element was 187 MYA and the youngest integration was recent. In the case of Dataset III, the structure was more complete in all datasets except 100kb downstream. The most ancient ERV in this dataset was integrated 190MYA and the newest only recently. Finally, in Dataset IV the range of insertion was 9-41.5 MYA.

The characteristics of the genes in overrepresented GO terms (Table V-6) were, in general, quite different from those of the genes around BoERVs (Table V-3). Predicted pseudogenes accounted for around 10% and the proportion of predicted genes varied between the different ranges and datasets. In addition, most of the genes within BoERVs were predicted (Table V-6) in all datasets.

Table V-6: Features of genes related with overrepresented GO terms in each range and BoERV dataset.

Range			Location					
	Included Gene in BoERV	BoERV in intron	Upstream (Kb)			Downstream (Kb)		
			5	50	100	5	50	100
Dataset I								
Genes	30.77%	-	80%	55%	56.7%	16.67%	54.63%	60.33%
Predicted genes	50%	-	20%	32.27%	28.49%	83.33%	34.94%	29.20%
Pseudogenes	0%	-	0%	0%	0%	0%	0%	0%
Predicted pseudogenes	19.23%	-	0%	12.73%	14.81%	0%	10.43%	10.47%
Dataset II								
Genes	30%	-	-	56%	-	-	43.33%	62.5%
Predicted genes	60%	-	-	44%	-	-	55%	37.5%
Pseudogenes	0%	-	-	0%	-	-	0%	0%
Predicted pseudogenes	10%	-	-	0%	-	-	1.67%	0%
Dataset III								
Genes	28.57%	-	-	65.5%	46.67%	-	33.33%	63.64%
Predicted genes	71.43%	-	-	37.5%	48.89%	-	66.67%	33.33%
Pseudogenes	0%	-	-	0%	0%	-	0%	0%
Predicted pseudogenes	0%	-	-	0%	4.44%	-	0%	3.03%
Dataset IV								
Genes	-	-	-	12.5%	59.37%	-	-	-
Predicted genes	-	-	-	87.5%	37.5%	-	-	-
Pseudogenes	-	-	-	0%	0%	-	-	-
Predicted pseudogenes	-	-	-	0%	3.13%	-	-	-

V.4. Discussion

In this study we report the characteristics of genes within and around bovine endogenous retroviruses (BoERVs). There have been various papers which have analysed the contribution of transposable elements (TE) or LTR retrotransposons to the genes of different species (Medstrand et al., 2002; Ganko et al., 2003; DeBarry et al., 2006; Lockton & Gaut, 2009), among them cow's genes (Almeida et al., 2007). In this study the analyses focus on BoERVs and the search is centred on BoERVs instead of genes.

The number of BoERV insertions is different in each range and dataset. However, as a general rule, most insertions happen in genes within BoERVs and BoERVs within genes, and the number of gene/BoERV associations decreases with the distance from BoERVs until it stabilises. As far as we know, the number of genes near a retroelement has only been examined in *C. elegans*, where in a range of up to 5kb 140 gene/retroelement associations were described, in a search performed with

124 LTR-retrotransposons (Ganko et al., 2003). However, there are more studies on the association of transposable elements (TEs) within genes. In those studies in humans and mice, the proportion of TEs in the UTR, intron or promoter regions of genes is found to be between 18% and 27% (Ganko et al., 2003; van de Lagemaat et al., 2003; Jordan et al., 2003). Here, we show that there are a great many BoERVs in gene introns. This appears to be a widespread feature in mammal genes.

In general, BoERVs do not show an integration orientation bias in regarding to transcription orientation. This observation agrees with work on *C. elegans* (Ganko et al., 2003), where LTR retrotransposons near genes show no integration bias. However, there are cases of insertion bias in some of the ranges and datasets analysed. In all datasets BoERVs within gene introns are inserted in an antisense orientation to a significant degree. This agrees with a previous paper (Almeida et al., 2007) in which cow's genes were analyzed to identify TEs and it was reported that in cow's genes LTR classes were preferentially inserted in an antisense orientation. Moreover, the antisense insertion of TEs has also been reported in human and mouse introns (van de Lagemaat et al., 2003) and LTR classes in humans (Medstrand et al., 2002). In addition, in a study of a resurrected human endogenous retrovirus from the HERV-K family (Brady et al., 2009), the retrovirus showed no insertion bias. It integrated randomly into the genes, though its endogenous counterparts showed an insertion bias (Brady et al., 2009). It has been suggested that the integration bias of ERVs in gene introns may be related to the survival of ERVs in intronic regions (van de Lagemaat et al., 2006). As a more likely explanation it has been proposed that sense insertion could cause premature termination of transcripts due to the LTR's polyadenylation signal and could thus be subject to negative selection (van de Lagemaat et al., 2006; Brady et al., 2009).

In addition, other cases of significant bias are found: BoERVs situated downstream are in an antisense orientation. As mentioned above, on the whole the ERVs within genes or at gene boundaries are in an antisense orientation, presumably in order to avoid interference from regulatory elements of ERVs in genes (Medstrand et al., 2002; Almeida et al., 2007).

The most surprising result is the analysis of overrepresented Gene Ontology terms in genes around BoERVs. It is intriguing to find that these overrepresented terms are mostly related to "defence against virus" and "chromatine assembly" terms.

This analysis has some limits: firstly, the detection of BoERVs is limited [Chapter II], as in other species. Thus, the analysis is limited to detectable elements and the BoERVs related to these overrepresented terms account for only a minor portion of all the BoERVs used and BoERVs that are near genes. Secondly, the characterisation of bovine genes in Gene Ontology (GO) is not complete, so there is a portion of genes that are not included in this analysis. Finally, some artifacts could be caused due to the used cow genome version and its annotation.

Despite these limitations, different BoERV datasets are used in an attempt to avoid bias, from wide sets, e.g. the "all detected ERVs" dataset to narrow sets, e.g. the dataset with BoERVs whose genes are detected by BLAST. Despite variability between different sets, terms such as "defence against virus" and "chromatine assembly" appear in each dataset. When the range is lengthened the number of genes increases, on the whole proportionally, but this increment does not entail an increase of terms. Thus, although more genes may be analysed, the same categories appear. Despite the limitation of information from GO and possible artifacts, the coverage of genes characterised in our data and the overall genes characterised from the cow genome are quite similar, so the bias of this information should be minimal.

The genome annotation is difficult to overcome, because some genes are bioinformatic predictions that need to be confirmed by experiment, and some are pseudogenes. Moreover, in the study of genes around LTR retrotransposons in *C. elegans* (Ganko et al., 2003), a similar proportion of predicted genes was detected around LTR retrotransposons. Thus, it seems that the appearance of this kind of gene in these analyses is difficult to avoid.

However, it is difficult to explain and to prove in what way the GO terms mentioned above could emerge in this analysis. A previous study has shown that genes with TE insertions in human and mouse are related to a defence response (van de Lagemaat et al., 2003), an observation that is reinforced by our data.

In the aforementioned study with a resurrected HERV (Brady et al., 2009), the resurrected HERV inserted preferentially into transcription units, in gene-rich regions and near features associated with active transcription units, whereas endogenous HERVs are located outside transcription units (Brady et al., 2009). The explanation suggested is that HERVs integrated first with targeting biases and then the proviruses that disrupted gene function were lost (Brady et al., 2009). Thus, the emergence of GO categories in our data should be interpreted as a result of this process.

Related to the appearance of these terms is knowledge of the mechanism of retroviral insertion, which has been studied but from which it is difficult to draw general conclusions (Wu & Burgess, 2004).

- Retroviral insertion near defence genes could interfere with their function, and insertion near histones could manipulate the transcription of genomic regions. It has also been reported that a retrovirus could activate the host DNA repair mechanism to facilitate its integration. Within this mechanism there is the chromatin remodelling process (Smith & Daniel, 2006).

- Another possibility could be a more opportunistic process. As mentioned above, a resurrected HERV inserted preferentially into transcription units (Brady et al., 2009), as well as HIV or MLV (Bushman et al., 2005). In *Drosophila* it has been suggested that the expression pattern of a region could affect the TE insertion probability (Fontanillas et al., 2007), that is, when a chromosomal region expresses the euchromatin is exposed and could become a target for TEs. Thus, in a retroviral infection, it is possible that a portion of ERVs could be inserted near “response to virus” genes because these genes are expressing. However, the opposite has also been suggested, i.e. that transcriptionally active DNA is not a preferred target for RAV-1 (an exogenous member of the avian sarcoma-leukosis virus genus) (Weidhaas et al., 2000) or is at least less favourable for integration (Maxfield et al., 2005).

- An alternative explanation for the preference could be the sequence or the structures related with the sequence. It has been suggested that certain retroviruses such as HIV or MLV (murine leukemia virus) have a preference for insertion in weak palindromic sequences (Wu et al., 2005) and that this preference could be related to DNA structure. In addition, retroviruses prefer the DNA in nucleosomes rather than naked DNA for insertion, due the bending of the target DNA (Muller & Varmus, 1994). Moreover, the P transposable element in *Drosophila* seems to recognise the structural features (A-phillycity, protein-induced deformability, B-DNA twist, and bendability) at insertion sites, rather than a specific motif (Liao et al., 2000). Thus, genes related to overrepresented GO terms could have a structure that is recognizable for retroviruses.

- Finally, it has been proposed that retroviral insertion could be beneficial for the hosts (Jern & Coffin, 2008; Arnaud et al., 2008). For example, it could be useful for the host to have retroviral sequences near defence genes in order to co-operate in retroviral interference. Cases have been reported of interference between endogenous and exogenous retroviruses, for instance, in sheep (Spencer et al., 2003). In sheep the interference could happen in two ways: the retroviral receptor could be blocked by competition or the newly created retroviruses could carry a defective gene that renders the retrovirus non-functional (Spencer et al., 2003; Arnaud et al., 2008).

As pointed out above, a portion of the genes in overrepresented terms could be pseudogenes (e.g. in Dataset I in the “gene in BoERVs” range they account for 19.23%), non-functional sequences or only predicted genes (e.g. in Dataset I in the “gene in BoERVs” range they account for 50%), although their proportion is considerably smaller than in all BoERV/gene associations (e.g. in

Dataset I in the “gene in BoERVs” range they account for 30.59% and 75.59% respectively). It has been suggested that eukaryotic genomes are a battlefield between genes and TEs (Blikstad et al., 2008; Maksakova et al., 2008) and a portion of these pseudogenes or predicted genes could be remains of TE/genome interactions.

Moreover, it is reported that ERVs could give plasticity to the genome in order to promote gene duplications and rearrangements in multigene families, such as the Major Histocompatibility Complex (Doxiadis et al., 2008). Thus, a portion of these pseudogenes or predicted genes could be remains of or attempts at events of this sort. The biggest proportions of pseudogenes or predicted genes are moreover detected in genes within BoERVs, i.e. sequences that could be remains of processed pseudogenes or retrogenes. It has been suggested that in humans and mice (Sakai et al., 2007) processed pseudogenes may have contributed more to the genome than duplicated genes, although few genes have played biologically important functions and both processes have produced a great deal of “junk”. Thus, it could be reasonable to assume that BoERV/gene associations in which genes are within BoERVs could be copies of processed pseudogenes. In addition, the BoERVs from this range involved in the overrepresentation keep their structure (65.38% in Dataset I, 77.78 in Dataset II and 66.67% in Dataset III; there are no overrepresented terms in genes within BoERVs in Dataset IV) and these processed pseudogenes are related to “defence against virus”. As mentioned above, the major histocompatibility complex, which is involved in the defence response, is an example of gene duplication mediated by ERVs (Doxiadis et al., 2008). Moreover, it has been reported that the human gene FAM8A1 (family with sequence similarity 8), an autosomal highly conserved protein, is transduced by a human ERV to make different pseudogenes (Jamain et al., 2001), that is, the ERV captures this family’s pseudogene sequence and makes different copies by retrotransposition. Thus, ERVs could be a source of duplications, although a large proportion of the attempts do not give rise to a functional sequence.

Otherwise the human IRGM gene, a member of the IRG (immunity related GTPases) family, which is related to resistance to pathogens, could have recovered its functionality due an insertion of an ERV (Bekpen et al., 2009). It seems that during the early stages of primate evolution this gene became non-functional, and it recovered its functionality in an ape/human ancestor when an ERV inserted upstream of the gene and the LTR of the ERV served as a functional promoter of IRGM (Bekpen et al., 2009). Thus, it cannot be ruled out that permissiveness in insertion of ERVs around defence related genes could be connected with the promoter regions of ERV LTR and those genes expressed in conditions where ERV LTRs are active.

The insertion times of the BoERVs involved in overrepresented terms varies widely, so insertion could be a process that is maintained over time. Indeed, there could be an overestimation of the most ancient elements due to the methodology used, since LTR divergence could be saturated. However, within the range, it seems that a large proportion of the BoERVs related to overrepresented terms inserted at 20-50 MYA, a range that overlaps the process of speciation of Ruminantia and Bovinae (Fernandez & Vrba, 2005). In addition, these BoERVs are just a portion of the total BoERVs used (e.g. in Dataset I in the “genes within BoERVs” range, they account for 0.19%) and BoERVs related to genes (e.g. in Dataset I in the “genes within BoERVs” range, they account for 6.74%). Thus, the insertion preference for genes related to the defence response and histones could be a feature of just some BoERVs. As discussed above, each retrovirus seems to have different preferences (Wu & Burgess, 2004; Bushman et al., 2005) and it has been proposed, although there are exceptions, that retroviruses from the same genera have the same insertion preferences (Brady et al., 2009). Thus, the observations of this study are related to some BoERVs rather than all BoERVs that have invaded the cow genome.

In short, despite its limitations, this study is a step towards determining the genomic context of BoERVs and could be a landmark in understanding the dynamics of some BoERVs, especially their preferences for insertion into the cow's genome and the implications for the dynamics of the host-

genome.

V.5. References

- Al-Shahrour, F., P. Minguez, J. Tarraga, D. Montaner, E. Alloza, J. Vaquerizas, L. Conde, C. Blaschke, J. Vera, and J. Dopazo. 2006. BABELOMICS: a systems biology perspective in the functional annotation of genome-scale experiments. *Nucleic Acids Research* **34**:W472-W476.
- Almeida, L., I. Silva, W. Silva, J. Castro, P. Riggs, C. Carareto, and M. Amaral. 2007. The contribution of transposable elements to *Bos taurus* gene structure. *Gene* **390**:180-189.
- Arnaud, F., M. Varela, T. Spencer, and M. Palmarini. 2008. Coevolution of endogenous Betaretroviruses of sheep and their host. *Cellular and Molecular Life Sciences* **65**:3422-3432.
- Bekpen, C., T. Marques-Bonet, C. Alkan, F. Antonacci, M. Leogrande, M. Ventura, J. Kidd, P. Siswara, J. Howard, and E. Eichler. 2009. Death and Resurrection of the Human IRGM Gene. *Plos Genetics* **5**.
- Blikstad, V., F. Benachenhau, G. Sperber, and J. Blomberg. 2008. Evolution of human endogenous retroviral sequences: a conceptual account. *Cellular and Molecular Life Sciences* **65**:3348-3365.
- Brady, T., Y. Lee, K. Ronen, N. Malani, C. Berry, P. Bieniasz, and F. Bushman. 2009. Integration target site selection by a resurrected human endogenous retrovirus. *Genes & Development* **23**:633-642.
- Bushman, F., M. Lewinski, A. Ciuffi, S. Barr, J. Leipzig, S. Hannenhalli, and C. Hoffmann. 2005. Genome wide analysis of retroviral DNA integration. *Nature Reviews Microbiology* **3**:848-858.
- DeBarry, J., E. Ganko, E. McCarthy, and J. McDonald. 2006. The contribution of LTR retrotransposon sequences to gene evolution in *Mus musculus*. *Mol Biol Evol* **23**:479-481.
- Doxiadis, G., N. de Groot, and R. Bontrop. 2008. Impact of endogenous intronic retroviruses on major histocompatibility complex class II diversity and stability. *Journal of Virology* **82**:6667-6677.
- Elsik, C., R. Tellam, K. Worley, and Bovine Genome Sequencing Anal Cons. 2009. The Genome Sequence of Taurine Cattle: A Window to Ruminant Biology and Evolution. *Science* **324**:522-528.
- Fernandez, M. and E. Vrba. 2005. A complete estimate of the phylogenetic relationships in Ruminantia: a dated species-level supertree of the extant ruminants. *Biological Reviews* **80**:269-302.
- Fontanillas, P., D. Hartl, and M. Reuter. 2007. Genome organization and gene expression shape the transposable element distribution in the *Drosophila melanogaster* euchromatin. *Plos Genetics* **3**:2256-2267.
- Ganko, E., V. Bhattachajee, P. Schliekelman, and J. McDonald. 2003. Evidence for the contribution of LTR retrotransposons to *C-elegans* gene evolution. *Mol Biol Evol* **20**:1925-1931.
- Ganko, E., C. Greene, J. Lewis, V. Bhattacharjee, and J. McDonald. 2006. LTR retrotransposon-gene associations in *Drosophila melanogaster*. *Journal of Molecular Evolution* **62**:111-120.
- Jamain, S., M. Giron-dot, P. Leroy, M. Clergue, H. Quach, M. Fellous, and T. Bourgeron. 2001. Transduction of the human gene FAM8A1 by endogenous retrovirus during primate evolution. *Genomics* **78**:38-45.
- Jern, P. and J. Coffin. 2008. Effects of Retroviruses on Host Genome Function. *Annual Review of Genetics* **42**:709-732.
- Johnson, W. and J. Coffin. 1999. Constructing primate phylogenies from ancient retrovirus sequences. *Proceedings of the National Academy of Sciences of the United States of America* **96**:10254-10260.
- Jordan, I. K., I. B. Rogozin, G. V. Glazko, and E. V. Koonin. 2003. Origin of a substantial fraction of human regulatory sequences from transposable elements. *Trends in Genetics* **19**:68-72.
- Katz, R., K. Gravuer, and A. Skalka. 1998. A preferred target DNA structure for retroviral integrase in vitro. *Journal of Biological Chemistry* **273**:24190-24195.
- Liao, G., E. Rehm, and G. Rubin. 2000. Insertion site preferences of the P transposable element in *Drosophila melanogaster*. *Proceedings of the National Academy of Sciences of the United States of America* **97**:3347-3351.
- Lockton, S. and B. Gaut. 2009. The Contribution of Transposable Elements to Expressed Coding Sequence in *Arabidopsis thaliana*. *Journal of Molecular Evolution* **68**:80-89.
- Maksakova, I., D. Mager, and D. Reiss. 2008. Keeping active endogenous retroviral-like elements in check: the epigenetic perspective. *Cellular and Molecular Life Sciences* **65**:3329-3347.
- Maxfield, L., C. Fraize, and J. Coffin. 2005. Relationship between retroviral DNA-integration-site selection and host cell transcription. *Proceedings of the National Academy of Sciences of the United States of America* **102**:1436-1441.
- Medstrand, P., L. van de Lagemaat, and D. Mager. 2002. Retroelement distributions in the human genome: Variations associated with age and proximity to genes. *Genome Research* **12**:1483-1495.
- Muller, H. and H. Varmus. 1994. Dna Bending Creates Favored Sites for Retroviral Integration - An Explanation for Preferred Insertion Sites in Nucleosomes. *Embo Journal* **13**:4704-4714.
- Nekrutenko, A. and W. H. Li. 2001. Transposable elements are found in a large number of human protein-coding genes. *Trends in Genetics* **17**:619-621.
- Sakai, H., K. Koyanagi, T. Imanishi, T. Itoh, and T. Gojobori. 2007. Frequent emergence and functional resurrection of processed pseudogenes in the human and mouse genomes. *Gene* **389**:196-203.

- Smith, J. and R. Daniel.** 2006. Following the path of the virus: The exploitation of host DNA repair mechanisms by retroviruses. *Acs Chemical Biology* **1**:217-226.
- Spencer, T., M. Mura, C. Gray, P. Griebel, and M. Palmarini.** 2003. Receptor usage and fetal expression of ovine endogenous betaretroviruses: Implications for coevolution of endogenous and exogenous retroviruses. *Journal of Virology* **77**:749-753.
- Tristem, M.** 2000. Identification and characterization of novel human endogenous retrovirus families by phylogenetic screening of the Human Genome Mapping Project database. *Journal of Virology* **74**:3715-3730.
- van de Lagemaat, L., J. Landry, D. Mager, and P. Medstrand.** 2003. Transposable elements in mammals promote regulatory variation and diversification of genes with specialized functions. *Trends in Genetics* **19**:530-536.
- van de Lagemaat, L., P. Medstrand, and D. Mager.** 2006. Multiple effects govern endogenous retrovirus survival patterns in human gene introns. *Genome Biology* **7**.
- Vogt, V.** 1997. Retroviral virions and genomes, p. 343-435. *In* J. Coffin, S. Hughes, and H. Varmos (eds.), *Retroviruses*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Weidhaas, J., E. Angelichio, S. Fenner, and J. Coffin.** 2000. Relationship between retroviral DNA integration and gene expression. *Journal of Virology* **74**:8382-8389.
- Wu, X. and S. Burgess.** 2004. Integration target site selection for retroviruses and transposable elements. *Cellular and Molecular Life Sciences* **61**:2588-2596.
- Wu, X., Y. Li, B. Crise, S. Burgess, and D. Munroe.** 2005. Weak palindromic consensus sequences are a common feature found at the integration target sites of many retroviruses. *Journal of Virology* **79**:5211-5214.

Genes around bovine endogenous retroviruses (BoERVs) are related to host defense and histone functions

V.6. Appendix

Supplementary table V-1: Summary of Gene Ontology analysis by Babelomics in dataset I.

Detailed information of table V-4.

GO ontology and level (% of annotated genes in gene set - genome)	GO Term	Dataset vs genome	Adjusted p-value
Included - ERV in gene (425 genes)			
Biological process - level 3 (16.24% - 15.3%)	response to biotic stimulus (GO:0009607)	87.94%-12.06%	9.58535e-06
Biological process - level 3 (16.24% - 15.3%)	defense response (GO:0006952)	83.79%-16.21%	0.000369986
Biological process - level 4 (16% - 14.61%)	response to other organism (GO:0051707)	90.44%-9.56%	7.60433e-06
Biological process - level 5 (16% - 13.5%)	response to virus (GO:0009615)	95.55%-4.45%	6.37007e-07
Biological process - level 6 (12.71% - 12%)	translation (GO:0006412)	77.86%-22.14%	0.0112716
Included - Intron (1008 genes)			
-			
5kb upstream (728 genes)			
Biological process - level 3 (18.27% - 15.3%)	response to biotic stimulus (GO:0009607)	79.1%-20.9%	0.00872849
Biological process - level 4 (17.45% - 14.61%)	response to other organism (GO:0051707)	82.16%-17.84%	0.0183047
Biological process - level 5 (16.62% - 13.5%)	response to virus (GO:0009615)	90.36%-9.64%	0.00767005
50kb upstream (5438 genes)			
Biological process - level 3 (17.3% - 15.3%)	response to biotic stimulus (GO:0009607)	68.15%-31.85%	0.00148682
Biological process - level 3 (17.3% - 15.3%)	defense response (GO:0006952)	63.28%-36.72%	0.0307204
Biological process - level 4 (16.44% - 14.61%)	response to other organism (GO:0051707)	71.31%-28.69%	0.00217187
Biological process - level 5 (15.28% - 13.5%)	response to virus (GO:0009615)	83.53%-16.47%	1.3057e-05
Biological process - level 5 (15.28% - 13.5%)	macromolecule biosynthetic process (GO:0009059)	59.48%-40.52%	0.0427148
Biological process - level 6 (13.31% - 12%)	cellular component assembly (GO:0022607)	66.24%-33.76%	0.0137901
Biological process - level 6 (13.31% - 12%)	DNA packaging (GO:0006323)	67.9%-32.1%	0.0168829
Biological process - level 6 (13.31% - 12%)	translation (GO:0006412)	61.15%-38.85%	0.0226869
Biological process - level 6 (13.31% - 12%)	chromosome organization and biogenesis (sensu Eukaryota) (GO:0007001)	66.13%-33.87%	0.0226869
Biological process - level 7 (9.03% - 8.28%)	macromolecule complex assembly (GO:0065003)	67.81%-32.19%	0.00247884
Biological process - level 7 (9.03% - 8.28%)	establishment and/or maintenance of chromatin architecture (GO:0006325)	68.28%-31.72%	0.0117894
Biological process - level 8 (7.54% - 6.84%)	protein-DNA complex assembly (GO:0065004)	71.01%-28.99%	0.00370956
Biological process - level 8 (7.54% - 6.84%)	chromatin assembly or disassembly (GO:0006333)	68.7%-31.3%	0.0235812
Biological process - level 9 (3.37% - 2.63%)	chromatin assembly (GO:0031497)	68.89%-31.11%	0.0219784
100kb upstream (9074 genes)			
Biological process - level 3 (17.02% - 15.3%)	biosynthetic process (GO:0009058)	55.88%-44.12%	0.0273211
Biological process - level 4 (16.24% - 14.61%)	cellular biosynthetic process (GO:0044249)	56.53%-43.47%	0.034223
Biological process - level 5 (15.1% - 13.5%)	macromolecule biosynthetic process (GO:0009059)	61.07%-38.93%	0.000106372
Biological process - level 5 (15.1% - 13.5%)	response to virus (GO:0009615)	76.81%-23.19%	0.00334994
Biological process - level 6 (13.29% - 12%)	translation (GO:0006412)	63.1%-36.9%	3.37333e-05
Biological process - level 7 (8.83% - 8.28%)	macromolecule complex assembly (GO:0065003)	63.9%-36.1%	0.0305608
Biological process - level 8 (7.33% - 6.84%)	protein-DNA complex assembly (GO:0065004)	67.23%-32.77%	0.022203
5kb downstream (704 genes)			
Biological process - level 5 (14.35% - 13.5%)	response to virus (GO:0009615)	90.59%-9.41%	0.0223371
50kb downstream (5505 genes)			
Biological process - level 3 (16.37% - 15.3%)	response to biotic stimulus (GO:0009607)	66.69%-33.31%	0.0114099
Biological process - level 3 (16.37% - 15.3%)	cellular developmental process (GO:0048869)	58.73%-41.27%	0.0241156
Biological process - level 4 (15.53% - 14.61%)	organelle organization and biogenesis (GO:0006996)	64.2%-35.8%	0.00101902
Biological process - level 4 (15.53% - 14.61%)	response to other organism (GO:0051707)	68.64%-31.36%	0.0273026
Biological process - level 4 (15.53% - 14.61%)	cell differentiation (GO:0030154)	58.88%-41.12%	0.0315589
Biological process - level 5 (14.31% - 13.5%)	chromosome organization and biogenesis (GO:0051276)	72.85%-27.15%	1.54628e-06
Biological process - level 5 (14.31% - 13.5%)	response to virus (GO:0009615)	83.16%-16.84%	2.31723e-05
Biological process - level 5 (14.31% - 13.5%)	macromolecule biosynthetic process (GO:0009059)	60.06%-39.94%	0.0151389
Biological process - level 5 (14.31% - 13.5%)	DNA metabolic process (GO:0006259)	61.79%-38.21%	0.0199761
Biological process - level 5 (14.31% - 13.5%)	cell development (GO:0048468)	60.41%-39.59%	0.0199761
Biological process - level 6 (12.44% - 12%)	DNA packaging (GO:0006323)	74.29%-25.71%	8.00155e-07
Biological process - level 6 (12.44% - 12%)	chromosome organization and biogenesis (sensu Eukaryota) (GO:0007001)	73.57%-26.43%	8.00155e-07
Biological process - level 6 (12.44% - 12%)	cellular component assembly (GO:0022607)	70.08%-29.92%	1.27585e-05
Biological process - level 6 (12.44% - 12%)	translation (GO:0006412)	60.7%-39.3%	0.0423543
Biological process - level 7 (8.34% - 8.28%)	establishment and/or maintenance of chromatin architecture (GO:0006325)	74.85%-25.15%	4.92226e-07
Biological process - level 7 (8.34% - 8.28%)	macromolecule complex assembly (GO:0065003)	71.46%-28.54%	2.6101e-06
Biological process - level 8 (6.83% - 6.84%)	protein-DNA complex assembly (GO:0065004)	77.39%-22.61%	3.35648e-08
Biological process - level 8 (6.83% - 6.84%)	chromatin assembly or disassembly (GO:0006333)	76.88%-23.12%	3.54451e-08
Biological process - level 9 (3.27% - 2.63%)	chromatin assembly (GO:0031497)	75.39%-24.61%	3.30431e-07
100kb downstream (9090 genes)			
Biological process - level 3 (16.85% - 15.3%)	cellular developmental process (GO:0048869)	57.59%-42.41%	0.029508
Biological process - level 4 (16.13% - 14.61%)	cell differentiation (GO:0030154)	57.54%-42.46%	0.0403315
Biological process - level 4 (16.13% - 14.61%)	response to other organism (GO:0051707)	65.69%-34.31%	0.0403315
Biological process - level 5 (14.86% - 13.5%)	chromosome organization and biogenesis (GO:0051276)	67.97%-32.03%	0.000131912
Biological process - level 5 (14.86% - 13.5%)	response to virus (GO:0009615)	78.26%-21.74%	0.000617786
Biological process - level 5 (14.86% - 13.5%)	macromolecule biosynthetic process (GO:0009059)	57.91%-42.09%	0.0296376
Biological process - level 5 (14.86% - 13.5%)	cell development (GO:0048468)	58.5%-41.5%	0.0296376
Biological process - level 6 (13.03% - 12%)	chromosome organization and biogenesis (sensu Eukaryota) (GO:0007001)	68.69%-31.31%	6.94496e-05
Biological process - level 6 (13.03% - 12%)	DNA packaging (GO:0006323)	69.13%-30.87%	8.67348e-05
Biological process - level 6 (13.03% - 12%)	cellular component assembly (GO:0022607)	65.64%-34.36%	0.000384924
Biological process - level 6 (13.03% - 12%)	translation (GO:0006412)	59.08%-40.92%	0.0351924
Biological process - level 7 (8.95% - 8.28%)	establishment and/or maintenance of chromatin architecture (GO:0006325)	69.22%-30.78%	0.000136065
Biological process - level 7 (8.95% - 8.28%)	macromolecule complex assembly (GO:0065003)	66.41%-33.59%	0.000245523
Biological process - level 8 (7.37% - 6.84%)	protein-DNA complex assembly (GO:0065004)	72.32%-27.68%	4.25515e-06
Biological process - level 8 (7.37% - 6.84%)	chromatin assembly or disassembly (GO:0006333)	71.12%-28.88%	1.61293e-05
Biological process - level 9 (3.25% - 2.63%)	chromatin assembly (GO:0031497)	71.26%-28.74%	1.40075e-05

Supplementary table V-2: Summary of Gene Ontology analysis by Babelomics in dataset II.
Detailed information of table V-4.

GO ontology and level (% of annotated genes in gene set - genome)	GO Term	Dataset vs genome	Adjusted p-value
Included - ERV in gene (180 genes)			
Biological process - level 3 (16.11% - 15.3%)	response to biotic stimulus (GO:0009607)	89.67%-10.33%	0.00398542
Biological process - level 3 (16.11% - 15.3%)	defense response (GO:0006952)	87.03%-12.97%	0.00772683
Biological process - level 4 (16.11% - 14.61%)	response to other organism (GO:0051707)	92.37%-7.63%	0.0016596
Biological process - level 5 (16.11% - 13.5%)	response to virus (GO:0009615)	96.55%-3.45%	0.000435868
Biological process - level 8 (3.89% - 6.84%)	chromatin assembly or disassembly (GO:0006333)	93.46%-6.54%	0.0167884
Biological process - level 8 (3.89% - 6.84%)	protein-DNA complex assembly (GO:0065004)	93.64%-6.36%	0.0167884
Included - Intron (888 genes)			
5kb upstream (293 genes)			
50kb upstream (2392 genes)			
Biological process - level 3 (18.31% - 15.3%)	response to biotic stimulus (GO:0009607)	70.54%-29.46%	0.0135049
Biological process - level 4 (17.27% - 14.61%)	response to other organism (GO:0051707)	75.7%-24.3%	0.00310358
Biological process - level 5 (16.05% - 13.5%)	response to virus (GO:0009615)	83.51%-16.49%	0.0121711
100kb upstream (4269 genes)			
5kb downstream (288 genes)			
50kb downstream (2446 genes)			
Biological process - level 4 (16.31% - 14.61%)	organelle organization and biogenesis (GO:0006996)	65.21%-34.79%	0.0453951
Biological process - level 5 (15.17% - 13.5%)	chromosome organization and biogenesis (GO:0051276)	76.12%-23.88%	2.462e-05
Biological process - level 5 (15.17% - 13.5%)	response to virus (GO:0009615)	85.03%-14.97%	0.00110116
Biological process - level 6 (12.96% - 12%)	chromosome organization and biogenesis (sensu Eukaryota) (GO:0007001)	77.39%-22.61%	6.37369e-06
Biological process - level 6 (12.96% - 12%)	DNA packaging (GO:0006323)	77.36%-22.64%	2.29326e-05
Biological process - level 6 (12.96% - 12%)	cellular component assembly (GO:0022607)	74.39%-25.61%	2.29326e-05
Biological process - level 7 (8.91% - 8.28%)	establishment and/or maintenance of chromatin architecture (GO:0006325)	77.42%-22.58%	2.2687e-05
Biological process - level 7 (8.91% - 8.28%)	macromolecule complex assembly (GO:0065003)	74.92%-25.08%	2.2687e-05
Biological process - level 8 (7.24% - 6.84%)	chromatin assembly or disassembly (GO:0006333)	79.23%-20.77%	7.4981e-06
Biological process - level 8 (7.24% - 6.84%)	protein-DNA complex assembly (GO:0065004)	79.71%-20.29%	7.4981e-06
Biological process - level 9 (3.68% - 2.63%)	chromatin assembly (GO:0031497)	76.48%-23.52%	0.00015942
100kb downstream (4369 genes)			
Biological process - level 3 (17.28% - 15.3%)	cellular developmental process (GO:0048869)	61.09%-38.91%	0.00296606
Biological process - level 4 (16.64% - 14.61%)	cell differentiation (GO:0030154)	60.89%-39.11%	0.00931779
Biological process - level 4 (16.64% - 14.61%)	organelle organization and biogenesis (GO:0006996)	61.84%-38.16%	0.0443172
Biological process - level 5 (15.34% - 13.5%)	chromosome organization and biogenesis (GO:0051276)	73.55%-26.45%	2.57405e-06
Biological process - level 5 (15.34% - 13.5%)	response to virus (GO:0009615)	84.18%-15.82%	1.71271e-05
Biological process - level 5 (15.34% - 13.5%)	cell development (GO:0048468)	62.59%-37.41%	0.00241786
Biological process - level 6 (13.53% - 12%)	chromosome organization and biogenesis (sensu Eukaryota) (GO:0007001)	74.31%-25.69%	1.05342e-06
Biological process - level 6 (13.53% - 12%)	cellular component assembly (GO:0022607)	72.08%-27.92%	1.56731e-06
Biological process - level 6 (13.53% - 12%)	DNA packaging (GO:0006323)	73.98%-26.02%	6.29516e-06
Biological process - level 7 (9.48% - 8.28%)	macromolecule complex assembly (GO:0065003)	72.15%-27.85%	4.07893e-06
Biological process - level 7 (9.48% - 8.28%)	establishment and/or maintenance of chromatin architecture (GO:0006325)	73.7%-26.3%	1.1667e-05
Biological process - level 8 (7.97% - 6.84%)	protein-DNA complex assembly (GO:0065004)	76.5%-23.5%	1.02988e-06
Biological process - level 8 (7.97% - 6.84%)	chromatin assembly or disassembly (GO:0006333)	75.54%-24.46%	2.53601e-06
Biological process - level 9 (3.91% - 2.63%)	chromatin assembly (GO:0031497)	73.43%-26.57%	4.3258e-05

Genes around bovine endogenous retroviruses (BoERVs) are related to host defense and histone functions

Supplementary table V-3: Summary of Gene Ontology analysis by Babelomics in dataset III.

Detailed information of table V-4.

GO ontology and level (% of annotated genes in gene set - genome)	GO Term	Dataset vs genome	Adjusted p-value
Included - ERV in gene (42 genes)			
Biological process - level 3 (21.43% - 15.3%)	response to biotic stimulus (GO:0009607)	95.89%-4.11%	6.35279e-05
Biological process - level 3 (21.43% - 15.3%)	defense response (GO:0006952)	93.51%-6.49%	0.0033079
Biological process - level 3 (21.43% - 15.3%)	behavior (GO:0007610)	97.28%-2.72%	0.0309007
Biological process - level 4 (19.05% - 14.61%)	response to other organism (GO:0051707)	96.69%-3.31%	0.000972389
Biological process - level 5 (19.05% - 13.5%)	response to virus (GO:0009615)	98.78%-1.22%	3.85019e-05
Included - Intron (236 genes)			
-			
5kb upstream (100 genes)			
-			
50kb upstream (797 genes)			
Biological process - level 3 (16.19% - 15.3%)	defense response (GO:0006952)	77.87%-22.13%	0.00327771
Biological process - level 3 (16.19% - 15.3%)	response to biotic stimulus (GO:0009607)	79.6%-20.4%	0.00327771
Biological process - level 4 (15.56% - 14.61%)	response to other organism (GO:0051707)	82.5%-17.5%	0.0150532
Biological process - level 5 (14.43% - 13.5%)	response to virus (GO:0009615)	90.8%-9.2%	0.00555589
100kb upstream (1501 genes)			
Biological process - level 3 (17.65% - 15.3%)	defense response (GO:0006952)	72.92%-27.08%	0.00441543
Biological process - level 3 (17.65% - 15.3%)	response to biotic stimulus (GO:0009607)	71.69%-28.31%	0.0386125
Biological process - level 5 (15.99% - 13.5%)	response to virus (GO:0009615)	85.87%-14.13%	0.0211188
Biological process - level 5 (15.99% - 13.5%)	chromosome organization and biogenesis (GO:0051276)	72.88%-27.12%	0.0420535
Biological process - level 6 (13.46% - 12%)	DNA packaging (GO:0006323)	75.86%-24.14%	0.0193006
Biological process - level 6 (13.46% - 12%)	chromosome organization and biogenesis (sensu Eukaryota) (GO:0007001)	74.55%-25.45%	0.0193006
Biological process - level 7 (9.19% - 8.28%)	establishment and/or maintenance of chromatin architecture (GO:0006325)	76.05%-23.95%	0.0241623
Biological process - level 7 (9.19% - 8.28%)	macromolecule complex assembly (GO:0065003)	72.51%-27.49%	0.0396039
Biological process - level 8 (7.46% - 6.84%)	chromatin assembly or disassembly (GO:0006333)	79.14%-20.86%	0.00141176
Biological process - level 8 (7.46% - 6.84%)	protein-DNA complex assembly (GO:0065004)	79.62%-20.38%	0.00141176
Biological process - level 9 (4.13% - 2.63%)	chromatin assembly (GO:0031497)	75.53%-24.47%	0.0164131
5kb downstream (84 genes)			
-			
50kb downstream (773 genes)			
Biological process - level 5 (14.62% - 13.5%)	response to virus (GO:0009615)	89.59%-10.41%	0.0410523
100kb downstream (1429 genes)			
Biological process - level 3 (17.28% - 15.3%)	response to biotic stimulus (GO:0009607)	74.27%-25.73%	0.0136013
Biological process - level 3 (17.28% - 15.3%)	defense response (GO:0006952)	70.27%-29.73%	0.025116
Biological process - level 3 (17.28% - 15.3%)	chromosome segregation (GO:0007059)	89.67%-10.33%	0.025116
Biological process - level 4 (16.38% - 14.61%)	mitotic cell cycle (GO:0000278)	82.4%-17.6%	0.0127038
Biological process - level 4 (16.38% - 14.61%)	response to other organism (GO:0051707)	76.46%-23.54%	0.0340293
Biological process - level 7 (9.52% - 8.28%)	M phase of mitotic cell cycle (GO:0000087)	82.03%-17.97%	0.0346265
Biological process - level 8 (7.56% - 6.84%)	mitosis (GO:0007067)	82.6%-17.4%	0.0232046

Supplementary table V-4: Summary of Gene Ontology analysis by Babelomics in dataset IV.

Detailed information of table V-4.

GO ontology and level (% of annotated genes in gene set - genome)	GO Term	Dataset vs genome	Adjusted p-value
Included - ERV in gene (4 genes)			
-			
Included - Intron (24 genes)			
-			
5kb upstream (5 genes)			
-			
50kb upstream (87 genes)			
Biological process - level 3 (21.84% - 15.3%)	cellular developmental process (GO:0048869)	83.59%-16.41%	0.0047307
Biological process - level 4 (21.84% - 14.61%)	organelle organization and biogenesis (GO:0006996)	87.33%-12.67%	0.00587869
Biological process - level 4 (21.84% - 14.61%)	cell differentiation (GO:0030154)	82.94%-17.06%	0.00797354
Biological process - level 5 (20.69% - 13.5%)	chromosome organization and biogenesis (GO:0051276)	93.3%-6.7%	0.000103417
Biological process - level 5 (20.69% - 13.5%)	DNA metabolic process (GO:0006259)	87.62%-12.38%	0.00262703
Biological process - level 5 (20.69% - 13.5%)	cell development (GO:0048468)	85.72%-14.28%	0.00262703
Biological process - level 6 (19.54% - 12%)	DNA packaging (GO:0006323)	93.89%-6.11%	6.62942e-05
Biological process - level 6 (19.54% - 12%)	chromosome organization and biogenesis (sensu Eukaryota) (GO:0007001)	93.12%-6.88%	7.52495e-05
Biological process - level 6 (19.54% - 12%)	cellular component assembly (GO:0022607)	91.55%-8.45%	0.000210856
Biological process - level 7 (17.24% - 8.28%)	establishment and/or maintenance of chromatin architecture (GO:0006325)	92.33%-7.67%	0.00031431
Biological process - level 7 (17.24% - 8.28%)	macromolecule complex assembly (GO:0065003)	89.94%-10.06%	0.00104753
Biological process - level 8 (14.94% - 6.84%)	chromatin assembly or disassembly (GO:0006333)	93.09%-6.91%	5.86038e-05
Biological process - level 8 (14.94% - 6.84%)	protein-DNA complex assembly (GO:0065004)	93.27%-6.73%	5.86038e-05
Biological process - level 9 (10.34% - 2.63%)	chromatin assembly (GO:0031497)	89.75%-10.25%	0.00050396
100kb upstream (169 genes)			
Biological process - level 3 (24.26% - 15.3%)	cellular developmental process (GO:0048869)	83.38%-16.62%	7.25491e-07
Biological process - level 3 (24.26% - 15.3%)	cellular component organization and biogenesis (GO:0016043)	77.67%-22.33%	0.000148593
Biological process - level 4 (22.49% - 14.61%)	organelle organization and biogenesis (GO:0006996)	88.07%-11.93%	9.43135e-08
Biological process - level 4 (22.49% - 14.61%)	cell differentiation (GO:0030154)	83.78%-16.22%	4.33358e-07
Biological process - level 4 (22.49% - 14.61%)	nucleobase, nucleoside, nucleotide and nucleic acid metabolic process (GO:0006139)	69.88%-30.12%	0.00557195
Biological process - level 4 (22.49% - 14.61%)	biopolymer metabolic process (GO:0043283)	65.44%-34.56%	0.0272935
Biological process - level 5 (21.89% - 13.5%)	chromosome organization and biogenesis (GO:0051276)	93.56%-6.44%	1.84977e-11
Biological process - level 5 (21.89% - 13.5%)	DNA metabolic process (GO:0006259)	88.73%-11.27%	6.45347e-09
Biological process - level 5 (21.89% - 13.5%)	cell development (GO:0048468)	86.12%-13.88%	2.96171e-08
Biological process - level 6 (20.71% - 12%)	DNA packaging (GO:0006323)	94.12%-5.88%	5.79799e-12
Biological process - level 6 (20.71% - 12%)	chromosome organization and biogenesis (sensu Eukaryota) (GO:0007001)	93.37%-6.63%	1.55866e-11
Biological process - level 6 (20.71% - 12%)	cellular component assembly (GO:0022607)	91.85%-8.15%	2.01256e-10
Biological process - level 7 (16.57% - 8.28%)	establishment and/or maintenance of chromatin architecture (GO:0006325)	93.25%-6.75%	1.9764e-11
Biological process - level 7 (16.57% - 8.28%)	macromolecule complex assembly (GO:0065003)	91.12%-8.88%	5.0854e-10
Biological process - level 8 (15.38% - 6.84%)	chromatin assembly or disassembly (GO:0006333)	93.52%-6.48%	4.25568e-12
Biological process - level 8 (15.38% - 6.84%)	protein-DNA complex assembly (GO:0065004)	93.69%-6.31%	4.25568e-12
Biological process - level 9 (10.65% - 2.63%)	chromatin assembly (GO:0031497)	90.36%-9.64%	1.38585e-10
5kb downstream (8 genes)			
-			
50kb downstream (74 genes)			
-			
100kb downstream (137 genes)			
-			

Genes around bovine endogenous retroviruses (BoERVs) are related to host defense and histone functions

V-5. Supplementary Table: Genes which appeared in two datasets or more in a GO term and range.

In italics genes which appeared in three datasets, in bold in four. The definitions can be consulted in Entrez Gene (<http://www.ncbi.nlm.nih.gov/gene>)

	Gene in BoERV	50kb upstream	100kb upstream	5kb downstream	50kb downstream	100kb downstream
response to virus	LOC513706	BECN1	IFN-tau-c1		IFNW1	BNIP3L
	LOC617135	<i>IFN-tau-c1</i>	LOC515549		LOC513706	IFNG
	LOC618859	LOC515549	LOC617556		LOC523509	IFNW1
	LOC618974	LOC523509	LOC618801		LOC538589	LOC513706
	LOC619113	LOC540442	LOC618974		LOC540442	LOC523509
	LOC783912	LOC616977	LOC619113		LOC616272	LOC538589
		LOC617112	LOC781424		LOC616977	LOC540442
		LOC617135	LOC782273		LOC617112	LOC616272
		LOC617556	TP-1		<i>LOC618801</i>	LOC616977
		LOC618801			LOC618859	LOC617112
		LOC618943			LOC618985	LOC618801
		LOC618974			LOC619113	LOC618859
		LOC619092			<i>LOC781424</i>	LOC618943
		LOC619113			LOC781732	LOC618985
		LOC781424			LOC781778	LOC619092
		LOC782273			LOC782273	LOC781424
		LOC783912			TP-1	LOC781732
		TP-1				LOC781778
						LOC781815
						LOC782273
						LOC783912
						TP-1
chromatin assembly	<i>LOC521580</i>	<i>H2AFZ</i>			H3F3A	H2AFX
	<i>LOC521581</i>	<i>LOC505183</i>			LOC404073	H3F3A
	<i>LOC527388</i>	<i>LOC509275</i>			LOC509275	LOC404073
	<i>LOC529646</i>	<i>LOC518961</i>			LOC512612	LOC509275
	<i>LOC529647</i>	<i>LOC521580</i>			LOC518961	LOC512612
	<i>LOC540148</i>	<i>LOC521581</i>			LOC519069	LOC517481
	<i>LOC618824</i>	<i>LOC523214</i>			LOC521580	LOC518961
		<i>LOC524808</i>			LOC521581	LOC519069
		<i>LOC527388</i>			LOC523214	LOC520120
		<i>LOC529646</i>			LOC527388	LOC521580
		<i>LOC529647</i>			LOC529646	LOC521581
		<i>LOC540148</i>			LOC529647	LOC523214
		<i>LOC613926</i>			LOC539875	LOC527388
		<i>LOC618824</i>			LOC540148	LOC527645
		LOC787528			LOC613345	LOC528329
					LOC616868	LOC529277
					LOC616875	LOC529646
					LOC618824	LOC529647
					LOC781223	LOC538911
					LOC783767	LOC539875
					LOC788668	LOC540148
					LOC789369	LOC613345
					MGC133641	LOC614958
					MGC137823	LOC614970
					MGC142335	LOC616776
					TNP1	LOC616790
						LOC616800
						LOC616819
						LOC616868
						LOC616875
						LOC618824
						LOC781223
						LOC781566
						LOC782971
						LOC783767
						LOC788668
						LOC789369
						MGC128489
						MGC133641
						MGC137823
						MGC142335
						TNP1

**VI. Behiaren erretrobirus
endogenoen (BoERV)
adierazpena behiaren (*Bos taurus*)
EST-etan**

VI.1. Sarrera

Erretrobirus endogenoak (ERV, *endogenous retrovirus*) ostalariaren hozi-lerroan txertatu diren erretrobirus exogenoen fase probiralak dira eta belaunaldietan zehar genoma ostalarian txertatuta mantendu dira (Boeke & Stoye, 1997).

ERVak ornodunetan detektatuak izan dira eta euren egitura hiru gene nagusi aurkitzen dira: *gag*, erretrobirusaren egitura proteinak kodetzen dituen genea; *pol*, erretrobirusaren erreplikaziorako eta insertziorako beharrezkoa den makinaria entzimatikoa kodetzen duen genea; eta *env*, erretrobirusaren estalkiaren glikoproteinak kodetzen dituen genea. Gene hauek osotzen duten zati zentralaren alboetan LTRak (*long terminal repeats*) aurkitzen dira, zeinetan kontrol eskualdeak kokatzen diren, hala nola, promotoreak eta areagotzaileak (Griffiths, 2001; Gifford & Tristem, 2003).

ERVak, oro har, 3 klasetan sailkatzen dira, erretrobirus exogenoen generoekin duten erlazio filogenetikoaren arabera: I klasea, *Gammaretrovirus* eta *Epsilonretrovirus* generoekin erlazioa dutenak; II klasea, *Alpharetrovirus*, *Betaretrovirus*, *Deltaretrovirus* eta *Lentivirus* generoekin erlazionatutakoak; eta III klasea, *Spumavirusekin* erlazioa dutenak. Erretrobirus exogenoen ezaugarri patogenikoak direla eta, euren kide endogenoak ostalaritik ezabatzeke joera dago edota ERVek euren funtzionaltasuna galtzeko (Gifford & Tristem, 2003; Gifford et al., 2005).

Hala ere, animali espezie ezberdinetan ERVen adierazpena gertatzen dela ikusi da (Jern & Coffin, 2008). Adibidez, gizakian mikromatrize, *Northern blot* eta ESTen bilaketen bidez (Stauffer et al., 2004; Okahara et al., 2004; Seifarth et al., 2005), ardian in situ hibridazioen bidez (Spencer et al., 2003) edo txerrian RT-PCRen bidez (Klymiuk et al., 2006). Bestalde, ERVen adierazpen hau ehun eta garapen-fase ezberdinetan aldakorra dela ikusi da (Okahara et al., 2004). Honetaz gain, elementu transposakorren (TE, *transposable elements*) adierazpena, ERVak barne, *Drosophila* ere ikertua izan da. ERVetan ikusi zen bezala, TEen adierazpena ehun eta garapen faseraren araberkoa da eta TE familia askoren adierazpena gertatzen da (Deloger et al., 2009).

ERVen adierazpena ostalariarentzat onuragarria izan daitekeela aditzera eman da, ostalariari funtzio berriak eman baititzaioke, beti ere ERVen adierazpena kontrolpean gertatzen bada (Gifford & Tristem, 2003; Jern & Coffin, 2008). Onuragarriak diren funtzio berri hauen artean, enbrioia-aren garapena eta babesak aurki dezakegu, ERVen *env* genearen gaitasun immunosupresoreak direla eta (Spencer et al., 2003; Okahara et al., 2004; Gimenez et al., 2009). Honetaz gain, ERVen adierazpena euren kide exogenoen aurreko defentsan lagundu dezaketela aditzera eman da, gene erretrobiralen interferentziaren bidez, hala nola, oiloetan (Denesvre et al., 2003) eta ardietan (Spencer et al., 2003; Arnaud et al., 2008). Azkenik, ERVen LTRak kontrol eskualdeak diren heinean, ERV txertaketek geneen adierazpena kontrolatzen lagun dezakete, promotore, areagotzaile eta motzitsasketa alternatiborako gune berriak eman baititzakete (Jern & Coffin, 2008).

Eragin onuragarrietaz gain, erretrobirus exogenoen izaera patogenikoa dela eta, ERVen adierazpena eragin kaltegarriekin ere lotu da, hala nola, minbiziarekin eta gaixotasun autoimmuneekin (Gifford & Tristem, 2003; Jern & Coffin, 2008). Nolanahi ere, kasu hauetan, ERVen adierazpena gaixotasunaren jatorria edo ondorio den ez dago argi (Jern & Coffin, 2008). Bestalde, ERVak ehun osasuntsuetan adierazten direnez, ERVen adierazpen-aldaketak dira aztertu direnak (Stauffer et al., 2004). Gizakietan, minbizian (Stauffer et al., 2004; Frank et al., 2008), eskizofrenian eta desoreka bipolarretan (Frank et al., 2005) ERVen familia batzuek adierazpen-patroi aldatua azaldu dute edo ERVen adierazpena minbiziaren aurreko tratamenduaren arrakastarekin lotu da (Golan et al., 2008). Baita ere gaixotasun autoimmuneekin erlazionatuta egon daitezkeela proposatu da, adibiderik argienak esklerosi multiplean aurkituz (Perron et al., 2009). Gaixotasun honetan HERV-W familiako sekuentziak adierazten direla behatu da, esklerosi multiplearen garapenean parte hartuz, gizakiaren ERV honen *env* genearen adierazpenaren aurrean sortzen den erantzun immunearen ondorioz. (Perron et al., 2009). Bestalde, HERV-K ERVaren aktibazioak melanomaren garapenean eragina

duela ondorioztatu da (Serafino et al., 2009), ERV honen presentzian melanoma-jatorria duten zelulak egoera eranskorretik egoera ez-eranskorrera aldatzen baitira.

Gizakian ERVen adierazpena modu masiboan aztertzeko hiru metodo erabili dira bereziki: hibridazio-probak (Okahara et al., 2004), mikromatrizeak (Seifarth et al., 2005) eta EST (*expressed sequence tags*) datubaseetan bilaketak (Stauffer et al., 2004). Azken honek, ehun eta momentu jakin batean adierazten diren geneen irudia ematen duen heinean, ERVen adierazpenari buruzko informazio asko eman lezake. ESTetan bilaketak egiteko BLASTen oinarritutako metodoak erabili dira (Stauffer et al., 2004; Flockerzi et al., 2008) eta metodo konplexuagoak ere, hots, Markov-en ezkutuko ereduak erabiltzea proposatu da (Oja et al., 2007).

Behia (*Bos taurus*) animalia interesgarria da, alde batetik hausnarkarien artean sekuentziatzen lehena izan delako eta, bestetik, dituen inplikazio ekonomikoengatik (Elsik et al., 2009). Bestalde, orain dela gutxi, hainbat ERV detektatu dira guk egindako aurreko lan batean [II kapitulua]. Azkenik, behiaren EST datu-basea nahiko ondo hornituta dago, animalien artean, EST kopuruaren arabera, 4. kokatzen baita (http://www.ncbi.nlm.nih.gov/dbEST/dbEST_summary.html).

Hortaz, lan honen helburua behiaren ehunetan ERVen adierazpena EST datu-basearen bidez detektatzea da, ERVen eta ostalariaren arteko elkarrekintzari buruzko jakintza hobetu nahian.

VI.2. Materialak eta metodoak

Lan honetan, behiaren ERVen (BoERV) adierazpena aztertzeko, behiaren genomak burututako BLAST bidezko ERV bilaketan 3 gene erretrobiral nagusiak (*gag*, *pol* eta *env*) detektatu zitzaizkien BoERVak aukeratu ziren [II kapitulua]. Irizpide hau 93 BoERVek bete zuten, gehienak I klaseko BoERV1 eta BoERV3 familietakoak (VI-1. taula gehigarria).

VI.2.1. EST bilaketak

2008/08/28an eskuragarri zeuden behiaren EST sekuentzietan (1517053 EST) aukeratutako BoERVen *gag*, *pol* eta *env* geneen bilaketak egin ziren. Honetarako BLAST bilaketak NCBI-ko zerbitzaria erabiliz burutu ziren (<http://www.ncbi.nlm.nih.gov/genome/seq/BlastGen/BlastGen.cgi?taxid=9913>). Bilaketaren emaitzak XML formatuan eskuratu ziren eta PHPn idatzitako script-en bidez emaitzak ustiari eta sailkatu ziren. Analisiak egiteko bit-score > 200 zuten emaitzak kontutan hartu ziren, hortaz, antzekotasun-maila altuko emaitzak aztertu ziren; eta soilik ehuna edo organoa identifikatuta zeukaten ESTak eskuratu ziren, hots, elkartua (*pooled*) bezala identifikatuta zeuden ESTak baztertu ziren. BoERVen geneak detektatu zitzaizkien ESTak ehun edo organoaren arabera sailkatu ziren eta, ahal zenean, garapen-fasearen arabera. Baita ere EST bakoitza adierazten zen BoERVarekin eta BoERV genearekin lotu zen ere. EST sekuentzia zati batekin BoERV batek baino gehiagok erlazioa bazuen, BLAST emaitzetan oinarrituz, bat egiten zuen sekuentziarik luzeena zuen BoERVari egokitu zitzaion EST sekuentzia horren jatorria. Kontrol moduan GAPDH (glizeraldehido-3-fosfato dehidrogenasa) genearen bilaketak egin ziren azaldutako prozedura erabilita. Ehun/organoen arteko konparazioa errazteko, ehun bakoitzaren EST kopuruek sortu ditzaketen alborapenak saihesten baitira, Stauffer et al.-ek (Stauffer et al., 2004) erabilitako proportzioa erabili zen, lan honetan EST-P laburduraz adieraziko dena:

$$\frac{\text{positiboak diren EST kopurua}}{\text{EST kopuru totala}} * 10^5$$

VI.2.1.1. ERV familien erlazio intraespezifikoa

Adierazten diren ERV familien zuhaitz filogenetikoa egiteko, lan honetan aztertutako 93 BoERVen *pol* genearen RT eskualdearen sekuentziak, gizakien ERVen adierazpena aztertutako lanetatik (Okahara et al., 2004; Seifarth et al., 2005; Prusty et al., 2008) lortutako eskualde horren 20 sekuentzia eta txerrien (Klymiuk et al., 2006) ERVen RT eskualdeko 10 sekuentzia erabili ziren. Sekuentzia nukleotidikoak MAFFT 5.861 (Katoh et al., 2005) programa erabiliz (FFT-NS-i aukera)

lerrokatu ziren eta MEGA 3.1 (Kumar et al., 2004) programa erabiliz Neighbor-Joining zuhaitz bat eraiki zen (aukerak: 1000 bootstrap, p-distatzia, bikote-ezabapena).

VI.2.2. Testuinguru genomikoaren analisia adierazten diren ERVetan

Txertaketa-preferentziak ezagutu nahian, adierazten ziren BoERVetatik gertu zeuden geneak analizatu ziren. Horretarako, posizio eta hein ezberdinetara egon zitezkeen geneak bilatu ziren: i) gene barnean zeuden BoERVak; ii) BoERV barnean zeuden geneak, hots, BoERVen sekuentziak alde batera eta bestera zituzten geneak; iii) BoERVetatik 5, 50 eta 100 kilobasetara urgora zeuden geneak; eta iv) BoERVetatik 5, 50 eta 100 kilobasetara urbehera zeuden geneak.

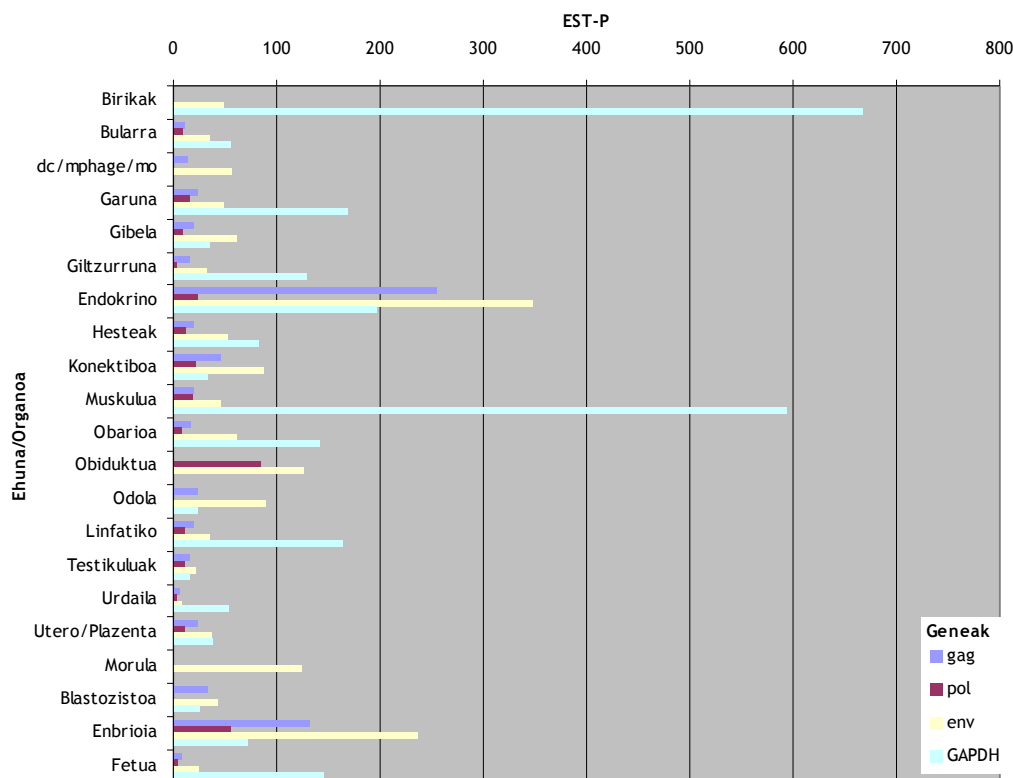
Adierazten ziren eta geneetatik gertu zeuden BoERVek txertaketa norabidean lehentasunik zuten aztertzeko χ^2 test bat burutu zen (Almeida et al., 2007) zorizko banaketa batekin alderatuta.

Adierazten ziren BoERVetatik gertu zeuden geneak Gene Ontology-ko termino gainerrepresentatuak bilatzeko erabili ziren. Horretarako Babelomics-eko (Al-Shahrour et al., 2006) FatiGO Compare lanabesa erabili zen (aukerak: gene zerrenda gainontzeko genomarena aurka; GO prozesu biologikoak databasea aukera lehenetsiekin; Fisherren test zehatza). Soilik doitutako p-balioak $< 0,05$ zirenak kontutan hartu ziren. Doitutako p-balioak FatiGO-k FDR (*false discovery rate*, aurkikuntza faltsuen tasa) prozedura erabiliz kalkulatu zituen (Al-Shahrour et al., 2006).

Gainerrepresentatutako terminoekin erlazioatutako balizko geneen inguruan zeuden BoERVen insertzio-denbora estimatzeko LTRen arteko dibergentzia erabili zen, hots, LTR arteko dibergentziari $2.3\text{-}5.0 \times 10^{-9}$ ordezen/leku/urte-ko tasa aplikatu zitzaion (Johnson & Coffin, 1999) eta, horrela, insertzio-denbora estimatu zen.

VI.3. Emaitzak

Aukeratutako behiaren erretrovirus endogenoen (BoERV) hiru genetarik bat gutxienez behiaren 16 ehun/organotatik eta 4 garapen-fasetatik lortutako ESTren baten detektatu zen (VI-1. irudia).



VI-1. irudia: gag, pol eta env ERVen geneen eta GAPDH genearen adierazpen erlatiboa aurkitututako ESTak ehun/organoka sailkatuta.

Adierazpen erlatiboa ((ESTak/ehunaren EST kopuru totala)* 10^5). GAPDH kontrol modura erabili da.

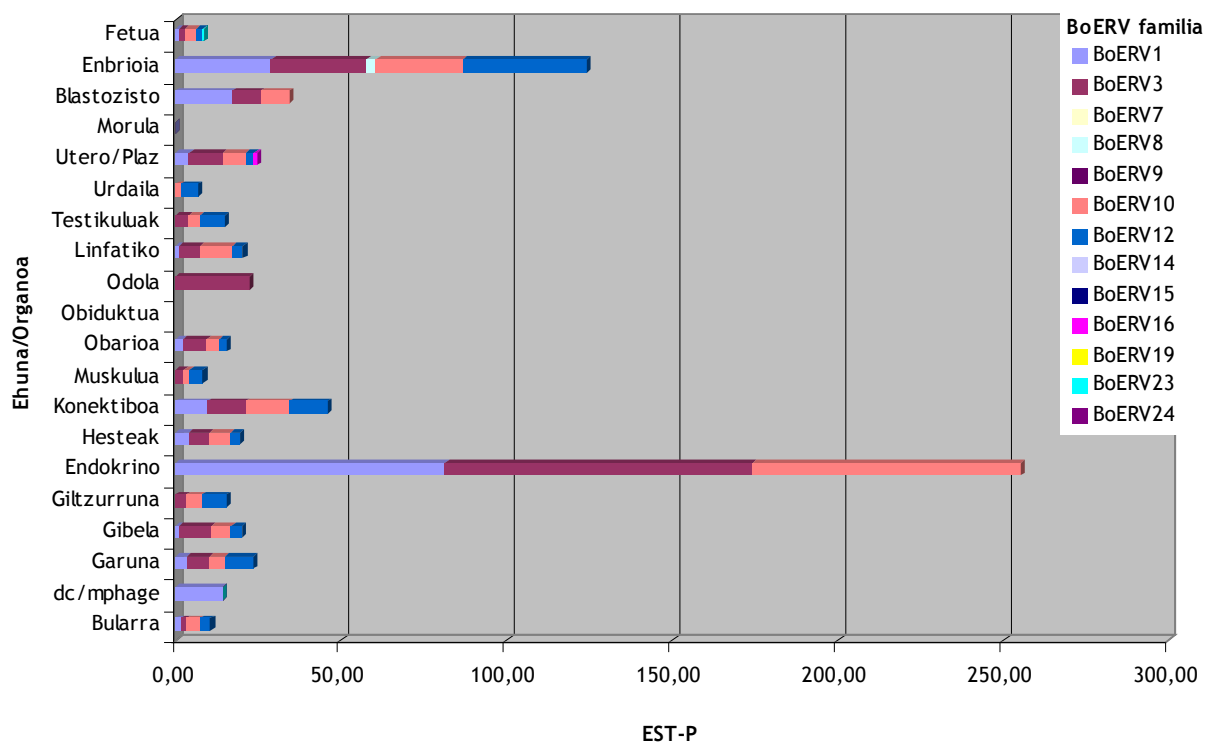
ESTen jatorrizko sailkapenak 16 ehun/organotan laburbildu ziren (VI-2. taula gehigarrian kontsultatu daitezkeenak) eta ehun/organo arteko konparazioa errazteko Stauffer et al-ek (Stauffer et al., 2004) erabilitako proportzioa erabili zen.

VI.3.1. Ehunetan adierazpena

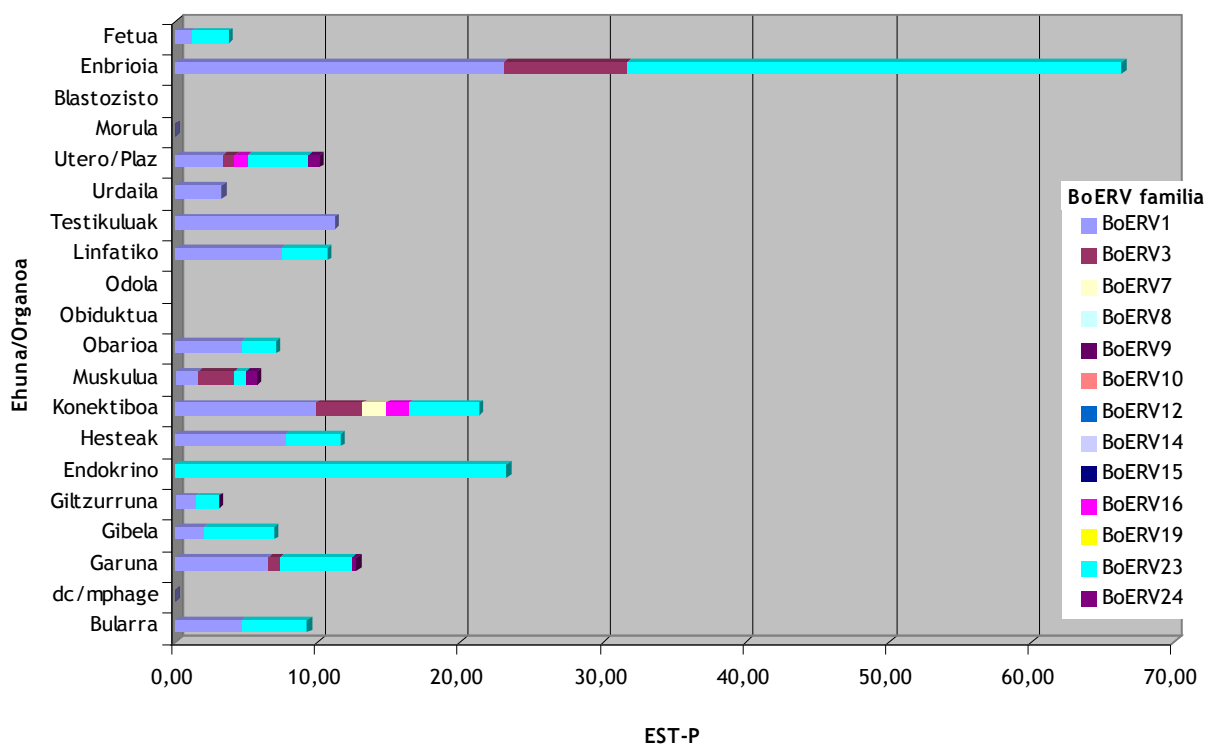
Oro har, aztertutako BoERVen adierazpen-mailarik altuena guruin endokrino eta enbrioietan detektatu zen eta baxuena urdailean (VI-1. irudia). Gainera, maila altuan adierazi ziren bi ehun horietan gene erretrobiralak GAPDH (glizeraldehido-3-fosfato dehidrogenasa), kontrol modura erabilitako genea, baino EST gehiagotan detektatu ziren. Adierazpen-maila ere gene erretrobiralaren araberrakoa izan zen, *env* genea ehun gehiagotan adierazi baitzen. Izan ere, *env* genea 16 ehun/organo eta 4 garapen-faseetan detektatu zen. Guruin endokrinoak, enbrioia eta morulak adierazpen mailarik altuenak azaldu zuten eta urdail, fetu eta testikulueta, aldiz, baxuenak. *pol* genearen kasuan, ordea, presentzia murriztuagoa azaldu zuen, 13 ehun/organotatik eta 2 garapen-faseetatik eratorritako ESTetan detektatu baitzen. *pol* genearen adierazpen mailarik altuena enbrioian detektatu zen eta baxuena giltzurrunetan. Bi gene hauen arteko adierazpena *gag* geneak izan zuen, 15 ehun/organotako eta 3 garapen-faseetako ESTetan aurkitu zena. *gag* sekuentzia gehien detektatu zen ehun/organoa guruin endokrinoa izan zen eta gutxien urdaila.

VI.3.2. BoERVen adierazpena familiaka

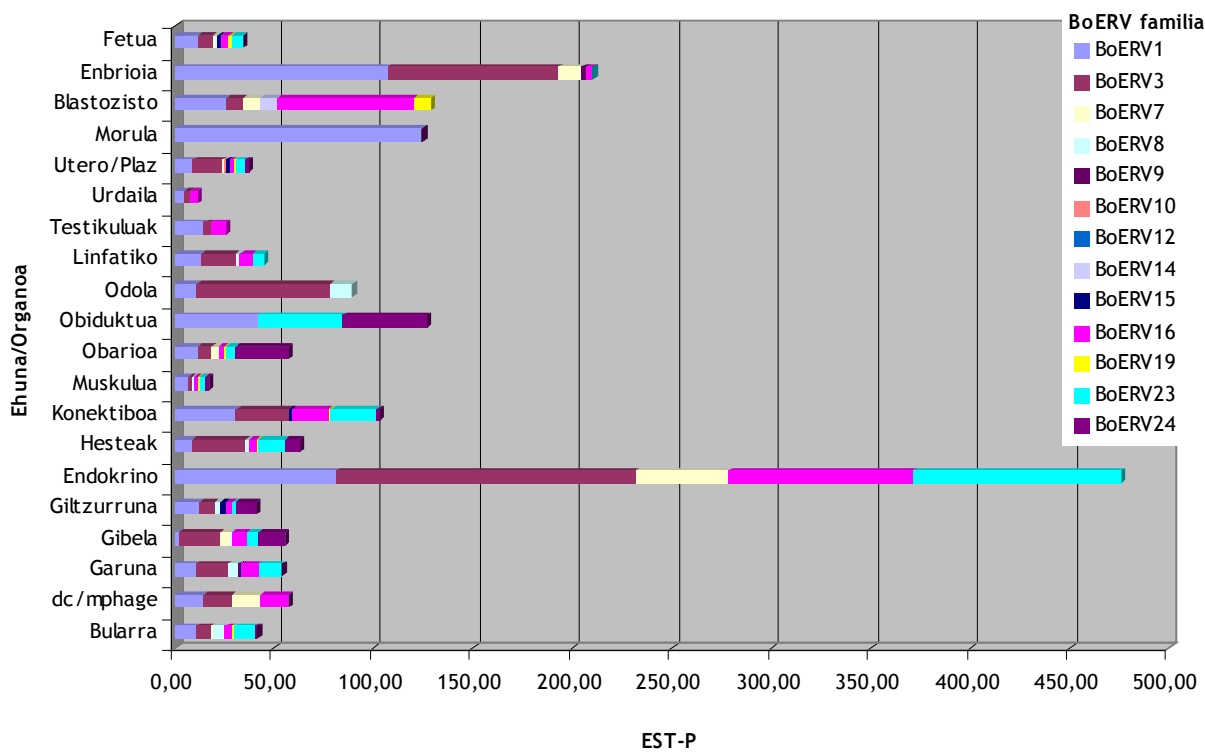
BoERV familia bakoitzaren kideen ekarpena ezberdina izan zen ehun, garapen-fase eta gene erretrobiralaren adierazpenaren ezberdintasunean. *gag* genearen kasuan (VI-2. irudia), BoERV1, BoERV3, BoERV10 eta BoERV12 familien kideak (I klasea) dira adierazpenaren jatorria. Euren adierazpena batez ere enbrioietan gertatzen da, eta adierazpena oso nabarmena da guruin endokrinoen kasuetan, BoERV12 familiaren kideetan izan ezik. BoERV1 familiarenak izan ezik, familia hauen kideak ere odolean detektatuak izan ziren. Maila baxuago batean lau familia hauen kideak garunean, hesteetan, gibelean, sistema linfatikoan eta obarioetan identifikatu ziren.



VI-2. irudia: *gag* genearen adierazpen-maila erlatiboa ((ESTak/ehunaren EST kopuru totala)*10⁵) ehun eta familiaka.



VI-3. irudia: *pol* genearen adierazpen maila erlatiboa ((ESTak/ehunaren EST kopuru totala)*10⁵) ehun eta familiak.



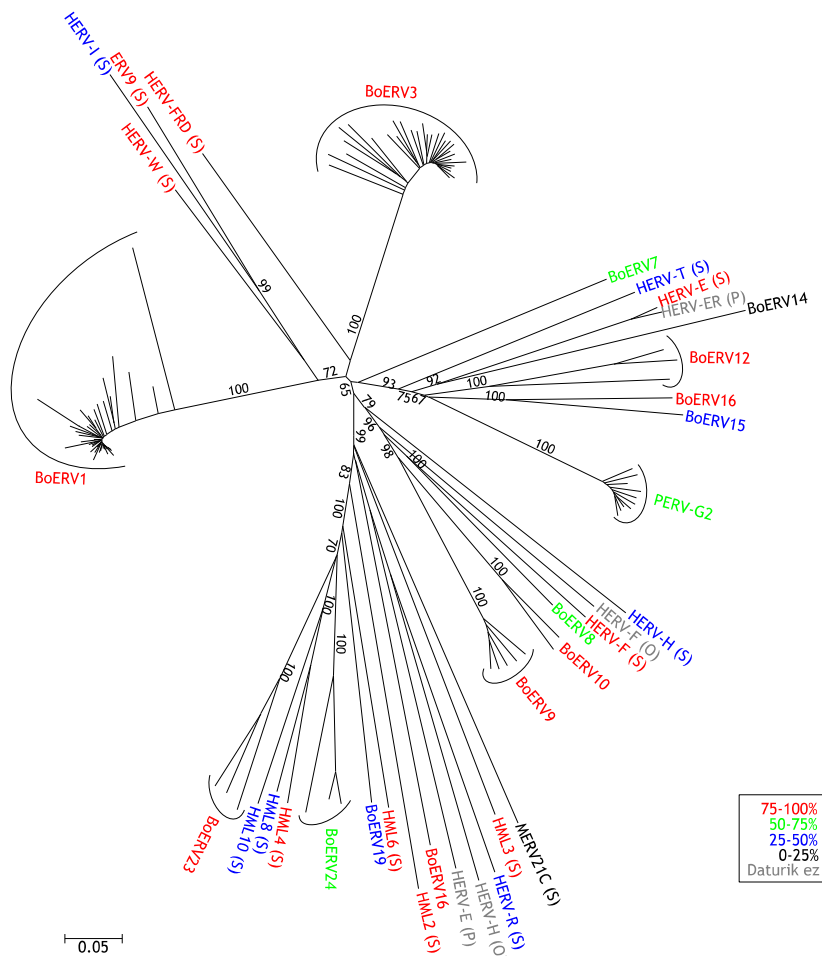
VI-4. irudia: *env* genearen adierazpen maila erlatiboa ((ESTak/ehunaren EST kopuru totala)*10⁵) ehun eta familiak.

pol geneari dagokionez (VI-3. irudia), BoERV1, BoERV3 (I klasea) eta BoERV23 (II klasea) familiako kideetan adierazten da batez ere. Hiru familia hauen kideen *pol* genearen adierazpena enbrioietan ematen da. Honetaz gain BoERV3 familiaren kideen *pol* genea ehun gutxitan adierazten da eta BoERV1 familiakoena garun, konektibo, heste, sistema linfatiko, bular eta barrabiletan. BoERV23 familiaren *pol* genea, aldiz, guruin endokrinotetan.

Azkenik, *env* genearen adierazpena (VI-4. irudia) BoERV1, BoERV3, BoERV16 (I klasea) eta BoERV23 (II klasea) familien kideetan zuen bere jatorria eta, modu apalago batean, BoERV7 eta BoERV24 familien kideetan. Familia hauen kideen *env* genea, BoERV24-koak izan ezik, guruin endokrinotetan nabarmenki adierazten da, batez ere BoERV3 familiakoena. Blastozistoen kasuan BoERV16 eta BoERV23 familien kideen adierazpena gertatzen da, batez ere BoERV23 familiarena eta enbrioia kasuan BoERV1, BoERV3 eta BoERV7 familien kideenak. Baita ere BoERV1 familiaren kideen *env* genea *in vitro* morulan nabarmenki adierazten da. Obiduktuaren kasuan BoERV1, BoERV23 eta BoERV24 familien kideen adierazpena gertatzen da eta obarioan, batez ere, BoERV24 familiarena. Azkenik BoERV3 familiaren kideak odolean ere adierazten dira.

VI.3.2.1. Adierazpena eta espezie ezberdinetako ERV familien arteko erlazioa

Lan honetan aztertutako BoERVen adierazpena eta beste mamaliotan adierazten diren ERV familia konparatzeko asmotan, adierazpena aztertutako behiaren, gizakiaren eta txerriaren ERVen familia ezberdinen ordezkarien zuhaitz filogenetikoa eraiki zen (VI-5. irudia). Zuhaitz horretan, I klasean barneratzen diren gizakiaren ERV9 eta HERV-W familiak behiaren BoERV1



VI-5. Irudia: Adierazpena aztertu zaien ERVen familien zuhaitz filogenetikoa.

NJ, 1000 bootstrap. Koloreetan, HERV/HML, gizakiaren ERVak; BoERV, behien ERVak; PERV, txerrien ERVak; familia bakoitzaren presentzia aztertutako ehun kopuruaren arabera ehunekotan adierazten da. Parentesi artean euren jatorria S: (Seifarth et al., 2005); P: (Prusty et al., 2008); O: (Okahara et al., 2004).

familiarekin multzokatu ziren, bi espezetan ehun askotan adierazpena duten familiak alegia. Era berean, I klaseko HERV-F eta BoERV9/BoERV10 familiak batera agertu ziren; BoERV8 familia, ordea, HERV-F-tik gertuago bazegoen ere beste bi behi familia baino ehun gutxiagotan adierazi zen. I klaseko gainontzeko familietan adierazten diren ehun kopuru eta ahaidetasun filogenetikoaren arteko joera hau ez zen hain argia izan. Adibidez, adierazten diren ehunen proportzio ezberdina zuten HERV-T/-E, BoERV12/14/15/16 eta PERV-G2 familiak batera multzokatu ziren. Hala ere, BoERV14/15 eta HERV-T familiak izan ezik, familia gehienak ehun askotan adierazpena azaldu zuten. Azkenik, II klasekoak diren gizakiaren HML4/8/10 eta behiaren BoERV19/23/24 familiek aztertutako ehunen laurdenean baino gehiagotan adierazi ziren.

VI.3.3. Testuinguru genomikoa

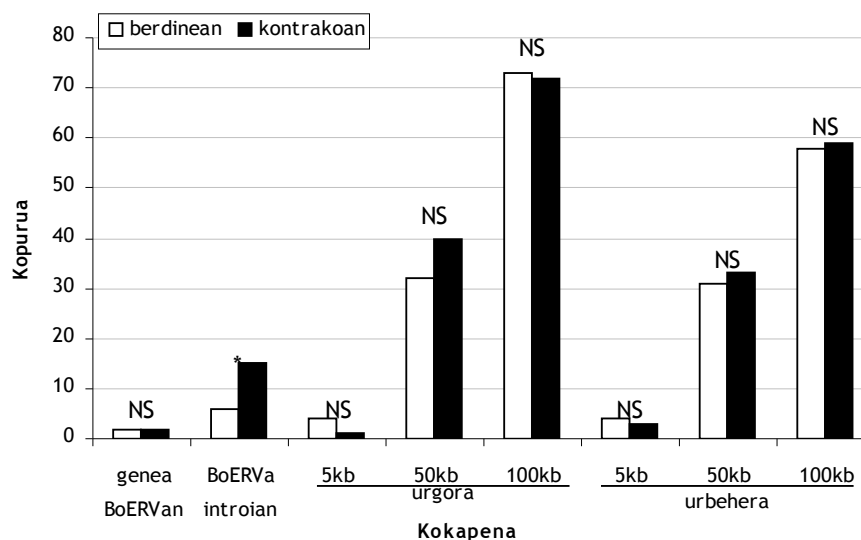
Adierazten ziren BoERVetatik gertu zeuden gene kopurua kokapen eta heinaren arabera izan zen (VI-1. taula). Itxarotekoa zenez, 5kb-ko heinak, bai urgora dauden geneetan (4 kasu) zein urbehera daudenetan (7 kasu), kasu gutxi aurkeztu zituen. 100Kb-etakoa, ordea kasu gehien (138 kasu urgoran eta 103 urbeheran).

Gene hauetatik auresandako geneen proportzioa (%25 -%100 artean) eta auresandako pseudogeneen proportzioa (%5-%25 artean) heinaren arabera aldakorrak izan ziren (VI-1. taula).

VI-1. taula: Adierazitako diren behiaren BoERVen inguruan dauden gene-kopurua.

Kokapena								
	Barneratuak		Urgora (Kb)			Urbehera		
Heina	Genea BoERVan	BoERVa introian	5	50	100	5	50	100
Gene/BoERV asoziazioak	4	21	5	72	145	7	64	117
ERVak (81 bakar)	4	21	5	72	145	7	64	117
Geneak (68 bakar)	4	20	5	40	51	6	39	57
Geneen ezaugarriak								
Aurresandakoak	50%	14.29%	0%	29.16%	20%	14.29%	29.69%	29.06%
Pseudogeneak	0%	0%	0%	1.38%	1.38%	0%	0%	0%
Aurresandako pseudogeneak	25%	19.06%	40%	12.5%	15.86%	14.29%	12.5%	13.68%

BoERVen txertaketa norabideari dagokionez (VI-6. irudia) ez zen txertaketa joera argirik egon, introietan agertzen diren BoERVetan izan ezik, zeintzuek transkripzio norabidearen aurka txertatzeko joera azaldu zuten.



VI-6. irudia: BoERVen txertaketa-norabidea.

NS, ez esangarria; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

BoERVetatik gertu zeuden geneak Gene Ontology-ko terminoren bat gainerrepresentatuta ote zegoen aztertzeko erabili ziren. Gainerrepresentatutako terminoak bakarrik 50kb eta 100kb urgora heinetan esangarriak izan ziren (VI-2. taula, VI-3. taula gehigarria) eta, oro har, kromosomen antolakuntzan parte hartzen duten geneak nabarmentzen dira.

Gainerrepresentatutako GO termino hauekin erlazonaturiko BoERVak I klasean barneratzen diren familietakoak izan ziren, batez ere BoERV1 eta BoERV3 familietakoak (VI-2. taula). BoERV hauen insertzio-denbora aldakorra izan zen. Bestalde, gainerrepresentatutako GO terminoekin erlazonatutako balizko geneen kopurua heinaren arabera aldakorra izan zen (7 eta 22 gene artean). Era berean, balizko geneen proportzio handi bat auresandakoa izan zen eta gene hauen proportzio oso txikia pseudogeneak izan ziren.

VI-2. taula: Gene Ontology termino gainerrepresentatuak eta erlazonatutako ERV eta geneak.

N: termino horrekin lotutako geneak; Esan.: esangarritasuna. Heina kb-etan.

	50kb		Kokapena Urgora	
	N	Esan.	N	Esan.
GO terminoak				
biopolymer metabolic process	-		20	*
nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	-		18	**
↳ DNA metabolic process	7	**	16	***
cellular component assembly	7	***	15	***
macromolecule complex assembly	7	***	15	***
↳ protein-DNA complex assembly	7	***	15	***
cellular component organization and biogenesis	-		16	***
↳ DNA packaging	7	*	15	***
↳ establishment and/or maintenance of chromatin architecture	7	***	15	***
↳ chromatin assembly or disassembly	7	***	15	***
↳ chromatin assembly	7	***	15	***
organelle organization and biogenesis	7	**	16	***
↳ chromosome organization and biogenesis	7	***	15	***
↳ chromosome organization and biogenesis (sensu Eukaryota)	7	***	15	***
cellular developmental process	7	*	16	***
↳ cell differentiation	7	*	16	***
↳ cell development	7	*	16	***
Erlazonatutako BoERVak (BoERV familia) [Insertzio-denbora (Orain dela milioi urte)]	23_14 (BoERV14) [57-26]		6_20 (BoERV1) [-]	
			4_19 (BoERV9) [33-15]	
			1_52 (BoERV1) [40-19]	
			21_18 (BoERV1) [0.434-0.2]	
			10_9 (BoERV3) [57-26]	
			18_18 (BoERV3) [13-6]	
			24_13 (BoERV3) [50-23]	
			1_23 (BoERV10) [29-13]	
			2_34 (BoERV16) [38-17]	
Erlazonatutako geneak	7		22	
Auresandakoak	100%		54.55%	
Pseudogeneak	0%		0%	
Auresandako pseudogeneak	0%		4.55%	

* p<0,05; ** p<0,01; *** p< 0,001

VI.4. Eztabaida

Lan honetan behien chunetan adierazi daitezkeen behien erretrovirus endogenoei (BoERV) buruzko lehen analisi sistematikoa aurkeztu da. Onartu beharra dago lan honetan erabilitako irizpideak eta metodologia direla eta, hemen aurkeztzen diren emaitzak benetan ematen den BoERVen adierazpenaren gutxiespena izan daitezkeela.

Alde batetik, oso kontserbatuak dauden BoERVak amu-sekuentzia bezala erabili dira, oinarritzko hiru gene erretrobiralak detektatu zaizkien BoERVak hain zuzen ere. Bestetik, BLAST bilaketak ez omen dira guztiz sakonak ESTetan erretrovirusak aurkitzeko (Oja et al., 2007; Flockerzi et al., 2008). Hala ere, metodo sofistikuagoak erabiltzen duen programarik ez dago eskuragarri eta orain arte mota honetako bilaketak egin diren lanak BLASTen oinarrituta daude (Stauffer et al., 2004; Deloger

et al., 2009). Bestalde, lan honetan erabilitako irizpideak zorrotzak izan dira, soilik antzekotasun-maila oso altua zuten emaitzak kontutan hartuz.

Oro har, BoERVak behiaren ehun ezberdinetan, maila ezberdinetara, adierazten dira eta, adierazpen hau BoERV familia bakoitzean ezberdina da. Emaiza hauek gizakian (Stauffer et al., 2004) eta *Drosophilan* (Deloger et al., 2009) egindako mota honetako lanetan ikusitakoaren ildo jarraitzen dute. *Drosophilaren* kasuan, elementu transposakorrak orokorrean aztertu ziren eta horien artean ERVekin erlazio urruna duten LTR erretrotransposoiak izan ziren adierazpen-mailarik altuenak azaldu zutenak (Deloger et al., 2009).

Ehunak/organoak aztertuz, guruin endokrinotan bai *gag* zein *env* geneak EST kopuru altuan detektatuak izan ziren, erabilitako kontrola (GAPDH) ere gaudituz. Adierazpen maila altu hau, batez ere, tiroideen eraginez gertatzen da (VI-2. taula gehigarria). Gizakiaren kasuan bai espermentalki (Shiroma et al., 2001) zein mikromatritzeen bidez (Seifarth et al., 2005) tiroideetan ERVen adierazpena ikusi da. Adierazpen hau familiaz familia aldakorra izan zen; batez ere HERV-E, HERV-W, HML2 eta HML-3 familietako kideak ziren adierazten zirenak (Seifarth et al., 2005). Hala ere euren funtzioa argitzeke dago.

Garapen-faseak aztertzerakoan, enbrioian BoERVen presentzia EST ugaritan detektatu zen. Batez ere *env* genea kontrolean baino EST gehiagotan agertu zen. ERVen *env* geneak ostalarietan, haurdunaldi garaian, onurak izan ditzakeela aditzera eman da, *env* genearen gaitasun fusogeniko eta inmunosupresoreak direla eta (Stoye & Coffin, 2000; Spencer et al., 2003). Honekin lotuta, nahiko deigarria da behiaren plazentaren ESTetan agertzen diren BoERVen kopuru maila baxuenetarikoa izatea. Ehun honetan ERVn adierazpena gizakietan, saguetan eta ardietan ikusi da, ERV baten *env* genea behar beharrezkoa delako plazentaren osaketarako, sinzitina hain zuzen ere (Jern & Coffin, 2008). Bai enbrioian zein plazentan gertatzen den ERVen *env* genearen adierazpena, enbrioaren babeserako gertatzen dela aditzera eman da (Stoye & Coffin, 2000; Spencer et al., 2003). Behiaren kasuan hau horrela izan daiteke, baina bakarrik enbrioian, orain arte aztertutako animalietan gertatzen denetik urrunduz. Gainera, gizakietan plazentan gertatzen den ERVen adierazpena plazentaren metilazio-maila txikiarekin lotuta egon daitekeela aditzera eman da (Okahara et al., 2004; Maksakova et al., 2008). Hortaz, hemen aurkeztzen diren emaitzek orain arte beste espezie batzuetan aditzera eman denetik aldentzen dira, behiaren plazentan BoERVek dinamika ezberdin bat daramatela aditzera emanez.

ERVen adierazpena modulatzeko duen mekanismo nagusia metilazioa da (Maksakova et al., 2008). Behiaren plazentan aztertutako gainontzeko espezieetan baino adierazpen gutxiago detektatzea azaldu ahal izateko behietan metilazio-patroia aztertua duten bi generen emaitzak aipa daitezke. Alde batetik, IGF2 (*insulin-like growth factor 2*) genearen inpronta-patroia dago. Gene hau behar beharrezko mitogeno fetala da eta behietan adin, ehun, promotore eta alelo arabera inpronta-patroi ezberdina azaldu zuen beste espezieekiko, hots, gizaki, sagu, txerri eta ardiekin (Curchoe et al., 2005). Bestetik, XIST (*X chromosome inactivation-specific transcript*) genearen inpronta-mekanismoa dago. Gene honen transkriptoak X kromosoma inaktibatzen du eta behi hibridotan (*Bos gaurus* x *Bos taurus*) gene honen eta alboetako eskualdeen inpronta-mekanismoa eta saguen mekanismoa ezberdinak direla ondorioztatu zuten (Dindot et al., 2004). Metilazioaz gain, behian ez da gizakian, saguan edo ardian (Jern & Coffin, 2008) bezala plazentaren morfogenesirako beharrezkoa den ERV espezifiko baten adierazpena deskribatu, sinzitina eta bere analogoak hain zuzen ere. Bestalde, “*Pregnancy-associated glycoproteins*” (PAG) deritzon peptidasa aspartikoen familia *Artiodactyla* ordeneko animalien plazentan bakarrik adierazten da. Gene hauek elementu transposakorren eragina jaso zutela aditzera eman bazen ere (Telugu et al., 2009), ERVen kasuan eragin hau nahiko txikia izan zen. Hala ere, BoERVak adierazten diren ESTetan PAG familiako kiderik modu sendoan detektatzeko ez gara gai izan. Hortaz, aurreko guztia uztartuz, beste espezieekin alderatuz behiaren plazentan adierazpen patroia eta dinamika ezberdinak ematen direla pentsa daiteke eta, hortaz, BoERVen *env* genearen adierazpena beste espezieetan bezalakoa ez izatea azaldu.

Metilazioarekin lotuta, *Arabidopsis*en TEekin egindako lan batean (Hollister & Gaut, 2009), TEak metilatzea gertu dauden geneetan eragina izan dezakeela aditzera ematen dute eta, horren ondorioz, hainbat TE ez direla metilatzen eragin kaltegarriak ekiditeko (Hollister & Gaut, 2009). Hortaz, lan honetan ikusten den BoERVen adierazpena ehun/organo batean momentu konkretu batean adierazten diren geneen funtzioa ez oztokatzearen ondorio izan daiteke.

pol genea ere enbrioian detektatua izan da, beste bi gene erretrobiralen maila berean izan ez bada ere. Hala ere, adierazpen hau nahiko interesgarria da. Saguen kasuan, enbrioia garapenaren hastapenetan, *pol* geneak kodetzen duen alderantzizko transkriptasa (RT, *reverse transcriptase*), lan honetan EST bilaketak egiteko erabili den eskualdea hain zuzen ere, enbrioia garapen normalarekin lotura duela ikusi da (Spadafora, 2004). RT entzimak zelularen garapen eta diferentziazioan zeregin aktiboa duela ikusi da, bai garapen normalean (enbriogenesian), zein aldatutako zelula-hazkundera (tumoretan) (Spadafora, 2004). Gainera, enbrioian RT entzimaren aktibitatea inhibitzen baldin bada, enbrioia garapena gelditzen da (Spadafora, 2004). Hortaz, lan honetan enbrioietan ikusitako *pol* genearen adierazpena enbrioian beharrezkoa dela adieraz lezake.

Adierazpena familiak aztertuz, genearen arabera, hainbat familiako kideek ehun guztietan adierazpena azaldu zuten, hala nola BoERV1, BoERV3 (I klasekoak) eta BoERV24 (II klasekoa) familietakoek. Familia hauen *pol* genearen adierazpena enbrioia ehunetan altua da. Interesgarriki, familia hauek behiaren genomaren gehien ugaritu diren familiak dira [III kapitulua]. Gizakien ERVen *pol* genearen adierazpena mikromatritze baten bidez aztertzen zuten lan batean (Seifarth et al., 2005), ugal ehunetan *pol* genearen adierazpena ERVen finkapenarekin erlazionatuta egon zitekeela aditzera eman zuten. Euren ustez, hozi-lerroko zeluletan ERVen transkripzio-maila altuak eta erretrotransposizioa beharrezkoak dira ERVen emendio eta finkapenerako eta, hortaz, ERVak ehun horietan adierazteko hautatzea (Seifarth et al., 2005). Agian gure emaitzetan ikusten den familia ugariaren *pol* adierazpena bide honetatik azaldu daiteke, hots, kide gehiago dituztela *pol* genearen aktibitatea enbriogenesian emendatuta dagoelako eta, ondorioz, beren buruaren kopia eta txertaketa berriak bultzatu direla.

Espezie ezberdinetan ehunez ehun aztertutako ERVen familien adierazpena eta familia hauen arteko erlazio filogenetikoa uztartzeko analisiak ez zuten emaitza argiegirik eman orokorrean. Nonahiko adierazpena duten familien arteko erlazio filogenetikoa egon daitekeela ondorioztatu daiteke. Honek, agian, etxekotu diren ERVekin lotuta egon daiteke, hots, eboluzioan zehar ostalarientzat, kasu honetan mamalioentzat, erabilgarri suertatu diren ERVen arteko erlazioari buruz. Lotura hau frogatzea zaila izateaz gain, ERV kopia bakoitzak metilazio patroia ezberdina duela aditzera eman da (Gimenez et al., 2009). Gorago aipatu bezala, metilazioa ERVen adierazpena kontrolatzeko mekanismo garrantzitsu bat da (Maksakova et al., 2008) eta gizakiaren ERVWE1 eta HERV-W familietan metilazio patroia familiekin erlaziorik ez duela ikusi da, metilazioa ERV kopia bakoitzaren LTR eta ehunarekin erlazioa duela baizik (Blanco et al., 2000).

Adierazten diren BoERVen inguruan dauden geneei dagokionez, ez zen txertaketa joerarik aurkitu, hots, zoriz transkripzioaren norabidean edo aurka txertatu ziren. Salbuespena introietan txertatzen ziren BoERVak izan ziren, kontrako norabidean txertatzeko joera azaldu zutenak. Behaketa hauek BoERV sorta ezberdinekin egindako lanaren ildo jarraitu zuten [V kapitulua].

Interesgarriki, gene hauen artean, BoERVetatik urgora dauden kasuan, kromosomen antolaketarekin erlazionatutako geneak, hots, adierazpena kontrolatzen duten geneak gainerrepresentatuta daude. Bestalde, emaitza hauek lehendik BoERVetatik gertu zeuden geneen azterketarekin [V kapitulua] bat datoz. Adierazten diren BoERVetatik gertu dauden geneek osotzen dituzten gainerrepresentatutako terminoak, hiru geneak detektatu zitzaizkien BoERVek osotzen zuten datu-multzoaren [V kapitulua] berdina izan ziren. Bat-etortze hau ez da guztiz harrigarria, adierazten diren BoERVak hiru geneak detektatu zitzaizkien BoERVen oso antzeko datu-multzoa baita. Honetaz gain, gainerrepresentatutako terminoekin erlazioa duten BoERVak, oro har, Ruminantia eta Bovinae taldeen espeziazioa gertatu zen garaioak dira (Fernandez & Vrba, 2005).

Laburbilduz, lan honetan aurkezten diren emaitzetan behien ehunetan hainbat BoERVren adierazpena detektatu da, nagusiki enbrioian eta guruin endokrinoetan. Emaitza hauek ERVen adierazpena behiaren ehun eta garapen-fasetan kontrolatuta dagoela aditzera ematen dute, beste espezie batzuetan detektatu den moduan.

VI.5. Bibliografia

- Al-Shahrour, F., P. Minguez, J. Tarraga, D. Montaner, E. Alloza, J. Vaquerizas, L. Conde, C. Blaschke, J. Vera, and J. Dopazo. 2006. BABELOMICS: a systems biology perspective in the functional annotation of genome-scale experiments. *Nucleic Acids Research* **34**:W472-W476.
- Arnaud, F., M. Varela, T. Spencer, and M. Palmarini. 2008. Coevolution of endogenous Betaretroviruses of sheep and their host. *Cellular and Molecular Life Sciences* **65**:3422-3432.
- Blanco, P., M. Shlumukova, C. A. Sargent, M. A. Jobling, N. Affara, and M. E. Hurles. 2000. Divergent outcomes of intrachromosomal recombination on the human Y chromosome: male infertility and recurrent polymorphism. *J Med Genet* **37**:752-758.
- Boeke, J. and J. Stoye. 1997. Retrotransposons, endogenous retroviruses, and the evolution of retroelements, p. 343-435. In J. Coffin, S. Hughes, and H. Varmos (eds.), *Retroviruses*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Blomberg, J., F. Benachenhov, V. Blikstad, G. r. Sperber, and J. Mayer. 2009. Classification and nomenclature of endogenous retroviral sequences (ERVs): Problems and recommendations. *Gene* **448**:115-123.
- Curchoe, C., S. Zhang, Y. Bin, X. Zhang, L. Yang, D. Feng, M. O'Neill, and X. Tian. 2005. Promoter-specific expression of the imprinted IGF2 gene in cattle (*Bos taurus*). *Biology of Reproduction* **73**:1275-1281.
- Deloger, M., F. M. G. Cavalli, E. Lerat, C. Biqmont, M. F. Sagot, and C. Vieira. 2009. Identification of expressed transposable element insertions in the sequenced genome of *Drosophila melanogaster*. *Gene* **439**:55-62.
- Denesvre, C., D. Soubieux, G. Pin, D. Hue, and G. Dambrine. 2003. Interference between avian endogenous ev/J 4.1 and exogenous ALV-J retroviral envelopes. *J Gen Virol* **84**:3233-3238.
- Dindot, S., K. Kent, B. Evers, N. Loskutoff, J. Womack, and J. Piedrahita. 2004. Conservation of genomic imprinting at the XIST, IGF2, and GTL2 loci in the bovine. *Mammalian Genome* **15**:966-974.
- Elsik, C., R. Tellam, K. Worley, and Bovine Genome Sequencing Anal Cons. 2009. The Genome Sequence of Taurine Cattle: A Window to Ruminant Biology and Evolution. *Science* **324**:522-528.
- Fernandez, M. and E. Vrba. 2005. A complete estimate of the phylogenetic relationships in Ruminantia: a dated species-level supertree of the extant ruminants. *Biological Reviews* **80**:269-302.
- Flockerzi, A., A. Ruggieri, O. Frank, M. Sauter, E. Maldener, B. Kopper, B. Wullich, W. Seifarth, N. Muller-Lantzsch, C. Leib-Mosch, E. Meese, and J. Mayer. 2008. Expression patterns of transcribed human endogenous retrovirus HERV-K(HML-2) loci in human tissues and the need for a HERV Transcriptome Project. *BMC Genomics* **9**:354.
- Frank, O., M. Giehl, C. Zheng, R. Hehlmann, C. Leib-Mosch, and W. Seifarth. 2005. Human endogenous retrovirus expression profiles in samples from brains of patients with schizophrenia and bipolar disorders. *Journal of Virology* **79**:10890-10901.
- Frank, O., C. Verbeke, N. Schwarz, J. Mayer, A. Fabarius, R. Hehlmann, C. Leib-Mosch, and W. Selfarth. 2008. Variable transcriptional activity of endogenous retroviruses in human breast cancer. *Journal of Virology* **82**:1808-1818.
- Gifford, R., P. Kabat, J. Martin, C. Lynch, and M. Tristem. 2005. Evolution and distribution of class II-related endogenous retroviruses. *Journal of Virology* **79**:6478-6486.
- Gifford, R. and M. Tristem. 2003. The evolution, distribution and diversity of endogenous retroviruses. *Virus Genes* **26**:291-316.
- Jimenez, J., C. Montgiraud, G. Oriol, J. P. Pichon, K. Ruel, V. Tsatsaris, P. Gerbaud, J. L. Frendo, D. Evain-Brion, and F. Mallet. 2009. Comparative Methylation of ERVWE1/Syncytin-1 and Other Human Endogenous Retrovirus LTRs in Placenta Tissues. *DNA Res* **16**:101-111.
- Golan, M., A. Hizi, J. Resau, N. Yaal-Hahoshen, H. Reichman, I. Keydar, and I. Tsarfaty. 2008. Human endogenous retrovirus (HERV-K) reverse transcriptase as a breast cancer prognostic marker. *Neoplasia* **10**:521-5U3.
- Griffiths, D. J. 2001. Endogenous retroviruses in the human genome sequence. *Genome Biol* **2**:REVIEWS1017.
- Hollister, J. D., B. S. Gaut. 2009. Epigenetic silencing of transposable elements: a trade-off between reduced transposition and deleterious effects on neighboring gene expression. *Genome Res* **19**: 1419-1428.
- Jern, P. and J. Coffin. 2008. Effects of Retroviruses on Host Genome Function. *Annual Review of Genetics* **42**:709-732.
- Johnson, W. and J. Coffin. 1999. Constructing primate phylogenies from ancient retrovirus sequences. *Proceedings of the National Academy of Sciences of the United States of America* **96**:10254-10260.
- Katoh, K., K. Kuma, H. Toh, and T. Miyata. 2005. MAFFT version 5: improvement in accuracy of multiple

sequence alignment. *Nucleic Acids Research* **33**:511-518.

Klymiuk, N., M. Muller, G. Brem, and B. Aigner. 2006. Phylogeny, recombination and expression of porcine endogenous retrovirus {gamma}2 nucleotide sequences. *J Gen Virol* **87**:977-986.

Kumar, S., K. Tamura, and M. Nei. 2004. MEGA3: Integrated software for molecular evolutionary genetics analysis and sequence alignment. *Briefings in Bioinformatics* **5**:150-163.

Maksakova, I., D. Mager, and D. Reiss. 2008. Keeping active endogenous retroviral-like elements in check: the epigenetic perspective. *Cellular and Molecular Life Sciences* **65**:3329-3347.

Oja, M., J. Peltonen, J. Blomberg, and S. Kaski. 2007. Methods for estimating human endogenous retrovirus activities from EST databases. *Bmc Bioinformatics* **8**.

Okahara, G., S. Matsubara, T. Oda, J. Sugimoto, Y. Jinno, and F. Kanaya. 2004. Expression analyses of human endogenous retroviruses (HERVs): tissue-specific and developmental stage-dependent expression of HERVs. *Genomics* **84**:982-990.

Perron, H., C. Bernard, J. B. Bertrand, A. B. Lang, I. Popa, K. Sanhadji, and J. Portoukalian. 2009. Endogenous retroviral genes, Herpesviruses and gender in Multiple Sclerosis. *Journal of the Neurological Sciences* **286**: 65-72.

Prusty, B., H. zur Hausen, R. Schmidt, R. Kimmel, and E. de Villiers. 2008. Transcription of HERV-E and HERV-E-related sequences in malignant and non-malignant human haematopoietic cells. *Virology* **382**:37-45.

Seifarth, W., O. Frank, U. Zeilfelder, B. Spiess, A. Greenwood, R. Hehlmann, and C. Leib-Mosch. 2005. Comprehensive analysis of human endogenous retrovirus transcriptional activity in human tissues with a retrovirus-specific microarray. *Journal of Virology* **79**:341-352.

Serafino, A., E. Balestrieri, P. Pierimarchi, C. Matteucci, G. Moroni, E. Oricchio, G. Rasi, A. Mastino, C. Spadafora, E. Garaci, and P. S. Vallebona. 2009. The activation of human endogenous retrovirus K (HERV-K) is implicated in melanoma cell malignant transformation. *Experimental Cell Research* **315**:849-862.

Shiroma, T., J. Sugimoto, T. Oda, Y. Jinno, and F. Kanaya. 2001. Search for active endogenous retroviruses: identification and characterization of a HERV-E gene that is expressed in the pancreas and thyroid. *Journal of Human Genetics* **46**:619-625.

Spadafora, C. 2004. Endogenous reverse transcriptase: a mediator of cell proliferation and differentiation. *Cytogenetic and Genome Research* **105**:346-350.

Spencer, T., M. Mura, C. Gray, P. Griebel, and M. Palmarini. 2003. Receptor usage and fetal expression of ovine endogenous betaretroviruses: Implications for coevolution of endogenous and exogenous retroviruses. *Journal of Virology* **77**:749-753.

Stauffer, Y., G. Theiler, P. Sperisen, Y. Lebedev, and C. V. Jongeneel. 2004. Digital expression profiles of human endogenous retroviral families in normal and cancerous tissues. *Cancer Immun* **4**:2.

Stoye, J. and J. Coffin. 2000. Reproductive biology - A provirus put to work. *Nature* **403**:715-+.

Telugu, B., A. Walker, and J. Green. 2009. Characterization of the bovine pregnancy-associated glycoprotein gene family - analysis of gene sequences, regulatory regions within the promoter and expression of selected genes. *BMC Genomics* **10**:185.

VI.6. Eranskina

VI-1. taula gehigarria: Lan honetan erabilitako ERVak. ERVen familia, beren posizioa Btau_3.1 genoma-bertsioan, ERVen izendapena Blomberg et al.-en (2009) gomendioa jarraituta eta ERV jakinaren gene erretrobiral bakoitza detektatu den EST kopurua.

ERV	Kromosoma: Posizioa (frame)	Izendapena	gag	pol	env
I klasea					
BoERV1					
chr1_43	1: 119642767-119651367 (+)	Btau_3.1_C_chr1_119642767_+			
chr 1_52	1: 135862148-135870733 (+)	Btau_3.1_C_chr1_135862148_+		2	
chr 1_54	1: 137728046-137736386 (-)	Btau_3.1_C_chr1_137728046_-			
chr 1_58	1: 145500205-145507889 (+)	Btau_3.1_C_chr1_145500205_+			10
chr 2_5	2: 1856883-1864414 (-)	Btau_3.1_C_chr2_1856883_-	1		
chr 2_8	2: 6865738-6877980 (+)	Btau_3.1_C_chr2_6865738_+	1		2
chr 2_26	2: 69350995-69361048 (-)	Btau_3.1_C_chr2_69350995_-	1		5
chr 3_22	3: 27178385-27188347 (+)	Btau_3.1_C_chr3_27178385_+			1
chr 3_47	3: 110191444-110200042 (+)	Btau_3.1_C_chr3_110191444_+			1
chr 4_41	4: 98567342-98583598 (-)	Btau_3.1_C_chr4_98567342_-	1	3	9
chr 5_39	5: 100600441-100610458 (+)	Btau_3.1_C_chr5_100600441_+			
chr 6_20	6: 26969323-26976989 (+)	Btau_3.1_C_chr6_26969323_+	3		2
chr 6_46	6: 102487500-102497561 (-)	Btau_3.1_C_chr6_102487500_-			
chr 8_29	8: 70005866-70015922 (-)	Btau_3.1_C_chr8_70005866_-		1	1
chr 9_32	9: 69574930-69582265 (-)	Btau_3.1_C_chr9_69574930_-			16
chr 11_12	11: 17083506-17091659 (-)	Btau_3.1_C_chr11_17083506_-	1		
chr 11_18	11: 32522822-32530490 (-)	Btau_3.1_C_chr11_32522822_-			18
chr 11_31	11: 99683892-99691690 (+)	Btau_3.1_C_chr11_99683892_+		82	
chr 13_19	13: 73184078-73197136 (+)	Btau_3.1_C_chr13_73184078_+			4
chr 14_3	14: 22670025-22684435 (-)	Btau_3.1_C_chr14_22670025_-		5	6
chr 15_3	15: 8414194-8422163 (-)	Btau_3.1_C_chr15_8414194_-	54		5
chr 15_7	15: 16150462-16160524 (-)	Btau_3.1_C_chr15_16150462_-			2
chr 15_23	15: 73269853-73279895 (+)	Btau_3.1_C_chr15_73269853_+			
chr 16_7	16: 5193188-5202190 (+)	Btau_3.1_C_chr16_5193188_+	12		9
chr 16_19	16: 34953459-34961067 (-)	Btau_3.1_C_chr16_34953459_-	4	2	1
chr 18_28	18: 56550392-56560439 (-)	Btau_3.1_C_chr18_56550392_-			
chr 20_15	20: 64131019-64139495 (-)	Btau_3.1_C_chr20_64131019_-	1		
chr 21_18	21: 32980977-32988658 (-)	Btau_3.1_C_chr21_32980977_-			7
chr 23_18	23: 28157049-28166863 (-)	Btau_3.1_C_chr23_28157049_-			1
chr 28_7	28: 11681840-11691910 (-)	Btau_3.1_C_chr28_11681840_-			
chr 29_1	29: 2145420-2153099 (+)	Btau_3.1_C_chr29_2145420_+			5
chr 29_3	29: 21070351-21079392 (+)	Btau_3.1_C_chr29_21070351_+		1	5
chr X_7	X: 24850350-24863470 (+)	Btau_3.1_C_chrX_24850350_+			100
chr X_10	X: 27283530-27291179 (-)	Btau_3.1_C_chrX_27283530_-	1		3
chr X_21	X: 37547260-37557493 (-)	Btau_3.1_C_chrX_37547260_-		4	9
chr X_25	X: 46324322-46335252 (*)	Btau_3.1_C_chrX_46324322_*		2	7
chr X_52	X: 63057274-63065872 (-)	Btau_3.1_C_chrX_63057274_-	2		
chr X_57	X: 65103350-65111911 (-)	Btau_3.1_C_chrX_65103350_-		1	
chr X_58	X: 65241297-65252481 (-)	Btau_3.1_C_chrX_65241297_-			48
BoERV3					
chr 1_34	1: 109771581-109780162 (+)	Btau_3.1_C_chr1_109771581_+			
chr 2_6	2: 4046523-4055117 (+)	Btau_3.1_C_chr2_4046523_+			2
chr 2_15	2: 26960400-26968977 (-)	Btau_3.1_C_chr2_26960400_-		2	1
chr 2_35	2: 89644805-89653575 (-)	Btau_3.1_C_chr2_89644805_-	99		1
chr 3_5	3: 7448153-7457355 (+)	Btau_3.1_C_chr3_7448153_+			1
chr 3_8	3: 9911997-9920968 (-)	Btau_3.1_C_chr3_9911997_-		4	61
chr 3_44	3: 97256295-97264809 (+)	Btau_3.1_C_chr3_97256295_+		1	2
chr 4_4	4: 4646655-4656128 (-)	Btau_3.1_C_chr4_4646655_-		3	
chr 4_7	4: 5161199-5169814 (+)	Btau_3.1_C_chr4_5161199_+	1		2
chr 4_18	4: 27408868-27418035 (-)	Btau_3.1_C_chr4_27408868_-			7
chr 4_24	4: 40874131-40883012 (+)	Btau_3.1_C_chr4_40874131_+			
chr 4_44	4: 104979058-104989110 (+)	Btau_3.1_C_chr4_104979058_+	1	1	99
chr 5_15	5: 40720694-40729627 (-)	Btau_3.1_C_chr5_40720694_-			
chr 6_14	6: 14658519-14667134 (+)	Btau_3.1_C_chr6_14658519_+			6
chr 6_22	6: 37492803-37502044 (-)	Btau_3.1_C_chr6_37492803_-			68
chr 7_6	7: 7977590-7986217 (+)	Btau_3.1_C_chr7_7977590_+	1	1	11
chr 8_17	8: 42699574-42708193 (-)	Btau_3.1_C_chr8_42699574_-			
chr 10_9	10: 21892543-21901517 (+)	Btau_3.1_C_chr10_21892543_+		1	1
chr 12_20	12: 37293618-37302919 (+)	Btau_3.1_C_chr12_37293618_+		1	
chr 13_3	13: 22659349-22668215 (-)	Btau_3.1_C_chr13_22659349_-		1	1
chr 14_7	14: 39011102-39019899 (-)	Btau_3.1_C_chr14_39011102_-	1		
chr 17_4	17: 4795060-4804035 (-)	Btau_3.1_C_chr17_4795060_-	51	1	1
chr 18_6	18: 14734098-14749262 (-)	Btau_3.1_C_chr18_14734098_-	1		3
chr 18_18	18: 52092475-52101086 (-)	Btau_3.1_C_chr18_52092475_-	1		1
chr 24_13	24: 47897308-47905807 (-)	Btau_3.1_C_chr24_47897308_-			1
chr 24_16	24: 54784761-54792198 (-)	Btau_3.1_C_chr24_54784761_-		1	7
chr 26_7	26: 9163919-9172972 (+)	Btau_3.1_C_chr26_9163919_+			2
chr 26_18	26: 36537863-36547093 (-)	Btau_3.1_C_chr26_36537863_-			1
chr 28_27	28: 39654356-39663304 (+)	Btau_3.1_C_chr28_39654356_+			
chr X_40	X: 54904883-54913819 (-)	Btau_3.1_C_chrX_54904883_-		1	
BoERV7					
chr 19_19	19: 36194623-36204816 (+)	Btau_3.1_C_chr19_36194623_+		1	49
BoERV8					
chr 20_1	20: 8659053-8667636 (-)	Btau_3.1_C_chr20_8659053_-	1		47
BoERV9					
chr 1_5	1: 8625962-8635985 (-)	Btau_3.1_C_chr1_8625962_-			3
chr 4_19	4: 28611868-28622158 (+)	Btau_3.1_C_chr4_28611868_+	1		5
chr 4_26	4: 45855666-45866206 (-)	Btau_3.1_C_chr4_45855666_-			10
chr 16_21	16: 42819360-42830121 (*)	Btau_3.1_C_chr16_42819360_*			4
chr X_1	X: 3304565-3315196 (+)	Btau_3.1_C_chrX_3304565_+			2
BoERV10					
chr 1_23	1: 76864086-76874394 (+)	Btau_3.1_C_chr1_76864086_+	10	1	1
chr 19_27	19: 56940310-56953304 (-)	Btau_3.1_C_chr19_56940310_-	100	1	
BoERV12					
chr 9_14	9: 26746755-26757441 (-)	Btau_3.1_C_chr9_26746755_-	100		

Behiaren erretrovirus endogenoen (BoERV) adierazpena behiaren (Bos taurus) EST-etan

ERV	Kromosoma: Posizioa (frame)	Izendapena	<i>gag</i>	<i>pol</i>	<i>env</i>
chr 19_17	19: 34885943-34896788 (-)	Btau_3.1_C_chr19_34885943_-			1
chr 25_8	25: 15049437-15056495 (+)	Btau_3.1_C_chr25_15049437_+			
BoERV14					
chr 23_14	23: 27514449-27524308 (+)	Btau_3.1_C_chr23_27514449_+			1
BoERV15					
chr 23_16	23: 27654628-27664824 (+)	Btau_3.1_C_chr23_27654628_+	1		17
BoERV16					
chr 2_34	2: 88708191-88719610 (+)	Btau_3.1_C_chr2_88708191_+	1		66
chr 13_22	13: 77896629-77919633 (-)	Btau_3.1_C_chr13_77896629_-		5	24
II klasea					
BoERV19					
chr 4_20	4: 33488089-33501364 (+)	Btau_3.1_B_chr4_33488089_+	2	2	13
BoERV23					
chr 3_13	3: 14947098-14958371 (+)	Btau_3.1_B_chr3_14947098_+		3	96
chr 15_8	15: 16253468-16264348 (-)	Btau_3.1_B_chr15_16253468_-		92	65
chr 26_5	26: 8210107-8221177 (+)	Btau_3.1_B_chr26_8210107_+	2	1	18
BoERV24					
chr 5_26	5: 58139244-58149091 (+)	Btau_3.1_B_chr5_58139244_+	1	1	1
chr 12_31	12: 56017095-56026005 (-)	Btau_3.1_B_chr12_56017095_-	1	3	48
chr 15_18	15: 49302007-49327037 (+)	Btau_3.1_B_chr15_49302007_+			14
chr 15_19	15: 49527014-49551862 (+)	Btau_3.1_B_chr15_49527014_+		3	15

VI-2. taula gehigarria: ERVen geneen adierazpena aurkitutako ESTetan, ehun/organoka sailkatuta.
 N, kopurua; EST-P. (N/EST Guztira)*10⁵ (Stauffer et al., 2004)

Ehuna/Organoa	EST kopurua	gag+		pol+		env+		GADPH+	
		N	Prop.	N	Prop.	N	Prop.	N	Prop.
Garuna	258999	61	23,55	40	15,44	126	48,65	439	169,50
anterior brain stem	5701					3	52,62	10	175,41
basal ganglia	16297	12	73,63	8	49,09	16	98,18	17	104,31
Brain	32666	3	9,18	1	3,06	4	12,25	63	192,86
brain cortex	5564					2	35,95	42	754,85
brain tissue	25334	12	47,37	5	19,74	25	98,68	15	59,209
cerebellum brain	7286	5	68,62	2	27,45	6	82,35	64	878,4
cerebral cortex	18416	6	32,58	5	27,15	13	70,59	9	48,871
hippocampus	34044	12	35,25	9	26,44	24	70,50	7	20,562
hypothalamus	89216	6	6,73			19	21,30	156	174,86
posterior brain	6543	1	15,28					48	733,61
thalamus	17932	4	22,31	10	55,77	14	78,07	8	44,613
Odola	8963	2	22,31			8	89,26	2	22,31
Blood	937	2	213,45			2	213,45	1	106,72
peripheral blood mononuclear cells	8026					6	74,76	1	12,46
Connective	60948	28	45,94	13	21,33	53	86,96	21	34,46
adipose	950					2	210,53	1	105,26
adipose tissue	9171	2	21,81	3	32,71	2	21,81	1	10,904
cartilage	7911	1	12,64	1	12,64	3	37,92		
Skin	38686	24	62,04	9	23,26	41	105,98	12	31,019
subcutaneous adipose tissue	3006	1	33,27			2	66,53	7	232,87
subcutaneous fat tissue	1224					3	245,10		
dc/mphage/mo	6944	1	14,40			4	57,60		
Gurur endokrinok	8612	22	255,46	2	23,22	30	348,35	17	197,40
adrenal gland	3315					1	30,17	16	482,65
anterior pituitary	1597			1	62,62	4	250,47	1	62,617
pituitary	1071	1	93,37						
thyroid	2629	21	798,78	1	38,04	25	950,93		
Hesteak	77656	15	19,32	9	11,59	41	52,80	64	82,41
ascending colon	38714	7	18,08	9	23,25	18	46,49	17	43,912
colon/ileum	2592					1	38,58		
ileum	11938	7	58,64			17	142,40	42	351,82
rectum	5117					1	19,54	2	39,085
small intestine	17004					1	5,88		
small intestine mucosa	2291	1	43,65			3	130,95	3	130,95
Giltzurruna	65130	10	15,35	2	3,07	21	32,24	84	128,97
Gibela	101180	21	20,76	9	8,90	63	62,27	36	35,58
liver	91976	16	17,40	4	4,35	57	61,97	36	39,141
liver tissue	9204	5	54,32	5	54,32	6	65,19		
Birika	4041					2	49,49	27	668,15
Sistema linfatikoa	93693	19	20,28	10	10,67	33	35,22	154	164,37
lymph node	9421	4	42,46			5	53,07	14	148,6
lymph nodes	2459					2	81,33	2	81,334
peyers patch	3171			1	31,54			6	189,21
spleen	15648			2	12,78	4	25,56	19	121,42
tonsil	4607	4	86,82	2	43,41	4	86,82	7	151,94
thymus	58387	11	18,84	5	8,56	18	30,83	106	181,55
Bular	172749	18	10,42	16	9,26	60	34,73	95	54,99
lactating mammary gland	3241					1	30,85	2	61,709
mammary	1360					1	73,53		
mammary (late involution)	2811	4	142,30	5	177,87	6	213,45		
mammary gland	142522	11	7,72	10	7,02	43	30,17	78	54,728
mastitic mammary (bmaa combined)	4261	1	23,47			2	46,94		
pre-partum mammary	10425	1	9,59			3	28,78	14	134,29
teat canal epithelium	4419					1	22,63		
virgin mammary	3710	1	26,95	1	26,95	3	80,86	1	26,954
Muskulu	121697	25	20,54	23	18,90	57	46,84	723	594,10
cardiac muscle	5111			4	78,26			70	1369,6
epithelial muscle	19089	4	20,95	2	10,48	16	83,82	15	78,579
heart ventricular	27426	2	7,29			5	18,23	86	313,57
muscle	35677	16	44,85	14	39,24	25	70,07	189	529,75
skeletal muscle	18945	2	10,56	2	10,56	3	15,84	362	1910,8
smooth muscle	15449	1	6,47	1	6,47	8	51,78	1	6,4729
Obario	85254	15	17,59	7	8,21	52	60,99	121	141,93
corpus luteum	748			1	133,69			1	133,69
corpus luteum (cl)	640					2	312,50	2	312,5
day 3 corpus luteum	1734	2	115,34			3	173,01	2	115,34
day 14 cl prostaglandin f2-alpha	1876	2	106,61			6	319,83	3	159,91
day 14 cl from a non-pregnant animal	1474	1	67,84			1	67,84		
day 14 cl from a pregnant animal	2014	1	49,65			3	148,96	2	99,305
ovarian	6366	1	15,71	3	47,13	3	47,13	18	282,75
ovarian follicle early selected	552					4	724,64		
ovarian follicle recruited	2455	2	81,47	1	40,73	2	81,47	2	81,466
ovarian ipsilateral non ovulated	1403					1	71,28	1	71,276
ovarian non ovulated	4023			1	24,86	2	49,71	19	472,28
ovarian ovulated	1729	1	57,84			3	173,51	7	404,86
ovary	60240	5	8,30	1	1,66	22	36,52	64	106,24
Obiduktu	2365	0	0,00	2	84,57	3	126,85	0	0,00
day 3 oviduct	804					2	248,76		
day 0 oviduct post-lh surge	1134					1	88,18		
oviduct	427			2	468,38				
Urdaila	60886	4	6,57	2	3,28	5	8,21	33	54,20
rumen	59850	4	6,68	2	3,34	3	5,01	33	55,138
rumen omasum abomasum reticulum	348					1	287,36		
stomach-reticulum	688					1	145,35		
Testikuluak	26900	4	14,87	3	11,15	6	22,30	4	14,87
testes	5040	2	39,68	1	19,84	1	19,84	2	39,683
testis	21860	2	9,15	2	9,15	5	22,87	2	9,1491
Utero eta plazenta	118379	28	23,65	13	10,98	44	37,17	45	38,01
caruncular and inter-caruncular regions	4752					2	42,09	1	21,044

Behiaren erretrovirus endogenoen (BoERV) adierazpena behiaren (Bos taurus) EST-etan

Ehuna/Organoa	EST kopurua	gag+		pol+		env+		GADPH+	
		N	Prop.	N	Prop.	N	Prop.	N	Prop.
day 5 uterus	1684	1	59,38			2	110,80		
day 16 uterus from a pregnant animal	1805					1	102,99		
day 18 uterus from a pregnant animal	971			1	102,99				
endometrial	296					1	337,84		
mixture of uterus and placenta	6249	3	48,01			5	80,01	7	112,02
placenta	53351	8	15,00	3	5,62	17	31,86	25	46,859
uterus	49271	16	32,47	9	18,27	16	32,47	12	24,355
In vitro ekoiztutako trinkotu aurreko morula	1606					2	124,53		
Blastozistoa	11611	4	34,45			5	43,06	3	25,84
in vitro derived blastocysts day 8	1703	1	58,72					2	117,44
in vivo derived blastocysts day 8	8148	1	12,27			2	24,55	1	12,273
nuclear-transfer derived blastocysts d8	1760	2	113,64			3	170,45		
Enbrioia	34675	46	132,66	19	54,79	82	236,48	25	72,10
20 flushed preimplantation embryos	8971	39	434,73	15	167,21	62	691,12		
2-cell in vitro produced embryo	1631	1	61,31						
day 16 conceptus	1601					1	62,46	1	62,461
early elongating embryos (15 mm)	964	2	207,47	1	103,73	1	103,73		
embryo	448	1	223,21			1	223,21	2	446,43
embryo (day 36 and day 64) and	10220					3	29,35	7	68,493
embryo (day 36 and day 64) extra-embryonic	7691	1	13,00	1	13,00	1	13,00	2	26,004
embryonic	3149	2	63,51	2	63,51	13	412,83	13	412,83
Fetua	318869	26	8,15	15	4,70	80	25,09	465	145,83
fetal brain	50516	4	7,92	4	7,92	16	31,67	10	19,796
fetal cerebellum	37300	4	10,72	1	2,68	10	26,81	117	313,67
fetal cerebral cortex	27976	1	3,57	3	10,72	18	64,34	43	153,7
fetal liver	36451	1	2,74			4	10,97	15	41,151
fetal lung	41696	4	9,59	1	2,40	6	14,39	36	86,339
fetal medulla	19650	1	5,09	2	10,18	2	10,18	39	198,47
fetal pancreas	51560	1	1,94	2	3,88			19	36,85
fetal pons	32836	5	15,23	1	3,05	10	30,45	64	194,91
fetal spinal column	16154	5	30,95	1	6,19	11	68,09	16	99,047
foetal muscle	4730					3	63,42	106	2241

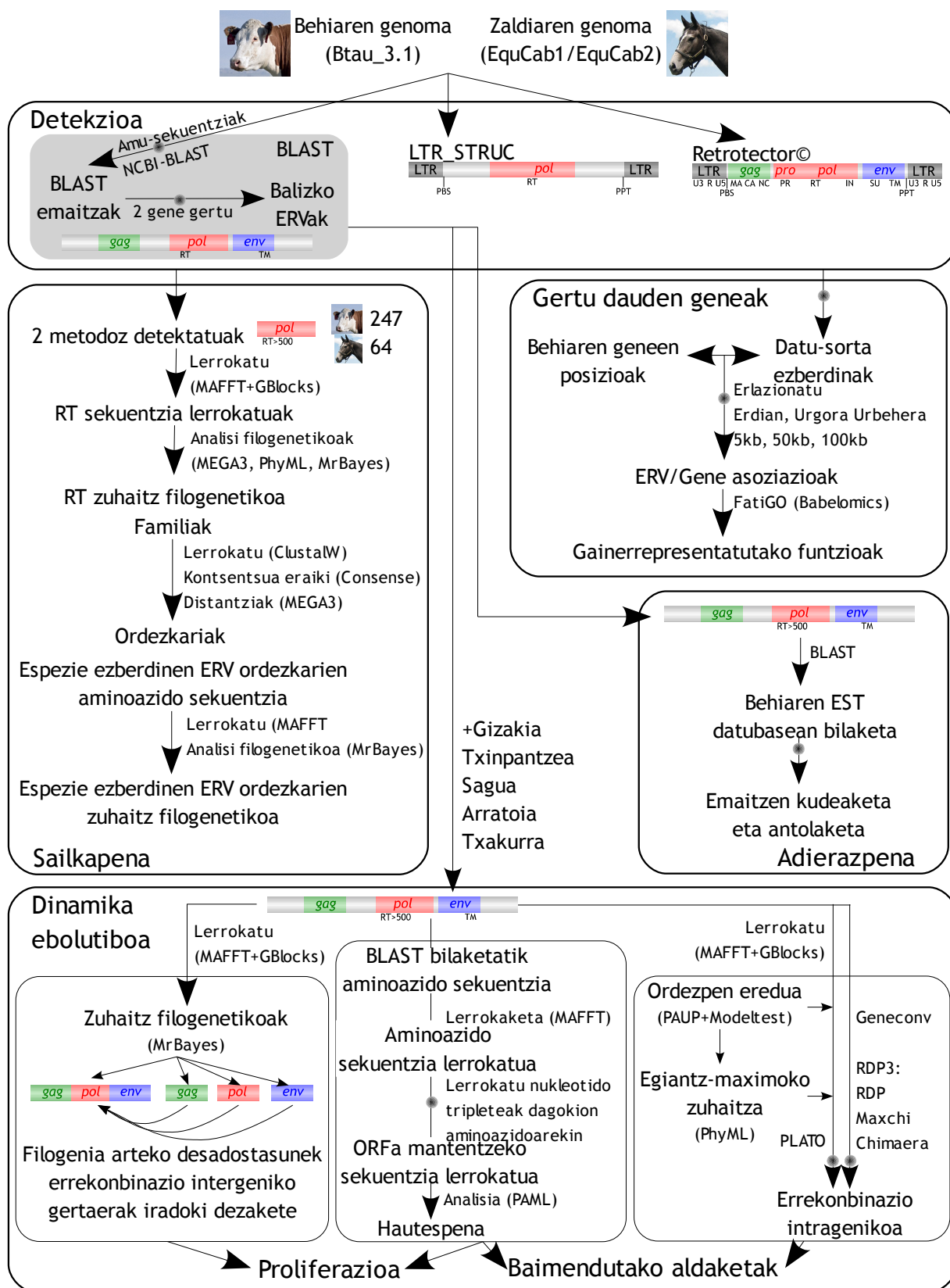
VI-3. taula gehigarria: Gene Ontology-ko emaitzak.

VI-1. taularen emaitza zehaztuak.

GO ontology and level % of annotated genes in gene set - genome	GO Term	Dataset vs genome	Adjusted p-value
Middle - ERV in gene (genes)			
Middle - Intron (genes)			
5000 bp upstream (genes)			
50000 bp upstream (72 genes)			
Biological process - level 3 (25% - 15.3%)	cellular developmental process (GO:0048869)	82.47%-17.53%	0.0239706
Biological process - level 4 (25% - 14.61%)	organelle organization and biogenesis (GO:0006996)	87.91%-12.09%	0.00390009
Biological process - level 4 (25% - 14.61%)	cell differentiation (GO:0030154)	81.79%-18.21%	0.0386103
Biological process - level 5 (23.61% - 13.5%)	chromosome organization and biogenesis (GO:0051276)	93.65%-6.35%	6.49015e-05
Biological process - level 5 (23.61% - 13.5%)	DNA metabolic process (GO:0006259)	88.23%-11.77%	0.00253338
Biological process - level 5 (23.61% - 13.5%)	cell development (GO:0048468)	84.76%-15.24%	0.0110178
Biological process - level 6 (22.22% - 12%)	DNA packaging (GO:0006323)	94.23%-5.77%	4.00175e-05
Biological process - level 6 (22.22% - 12%)	chromosome organization and biogenesis (sensu Eukaryota) (GO:0007001)	93.5%-6.5%	4.557e-05
Biological process - level 6 (22.22% - 12%)	cellular component assembly (GO:0022607)	92.01%-7.99%	0.000128559
Biological process - level 7 (19.44% - 8.28%)	establishment and/or maintenance of chromatin architecture (GO:0006325)	92.8%-7.2%	0.000174052
Biological process - level 7 (19.44% - 8.28%)	macromolecule complex assembly (GO:0065003)	90.55%-9.45%	0.000587106
Biological process - level 8 (18.06% - 6.84%)	chromatin assembly or disassembly (GO:0006333)	93.09%-6.91%	5.86038e-05
Biological process - level 8 (18.06% - 6.84%)	protein-DNA complex assembly (GO:0065004)	93.27%-6.73%	5.86038e-05
Biological process - level 9 (12.5% - 2.63%)	chromatin assembly (GO:0031497)	89.75%-10.25%	0.00050396
100000 bp upstream (145 genes)			
Biological process - level 3 (26.9% - 15.3%)	cellular developmental process (GO:0048869)	83.23%-16.77%	2.29616e-06
Biological process - level 3 (26.9% - 15.3%)	cellular component organization and biogenesis (GO:0016043)	78.52%-21.48%	6.82346e-05
Biological process - level 4 (24.83% - 14.61%)	organelle organization and biogenesis (GO:0006996)	88.63%-11.37%	3.78801e-08
Biological process - level 4 (24.83% - 14.61%)	cell differentiation (GO:0030154)	83.69%-16.31%	1.34075e-06
Biological process - level 4 (24.83% - 14.61%)	nucleobase, nucleoside, nucleotide and nucleic acid metabolic process (GO:0006139)	69.88%-30.12%	0.00836609
Biological process - level 4 (24.83% - 14.61%)	biopolymer metabolic process (GO:0043283)	65.56%-34.44%	0.0331269
Biological process - level 5 (24.14% - 13.5%)	chromosome organization and biogenesis (GO:0051276)	93.89%-6.11%	6.82137e-12
Biological process - level 5 (24.14% - 13.5%)	DNA metabolic process (GO:0006259)	89.27%-10.73%	2.28148e-09
Biological process - level 5 (24.14% - 13.5%)	cell development (GO:0048468)	86.06%-13.94%	1.0493e-07
Biological process - level 6 (22.76% - 12%)	DNA packaging (GO:0006323)	94.44%-5.56%	1.9664e-12
Biological process - level 6 (22.76% - 12%)	chromosome organization and biogenesis (sensu Eukaryota) (GO:0007001)	93.73%-6.27%	5.32299e-12
Biological process - level 6 (22.76% - 12%)	cellular component assembly (GO:0022607)	92.28%-7.72%	6.9738e-11
Biological process - level 7 (17.93% - 8.28%)	establishment and/or maintenance of chromatin architecture (GO:0006325)	93.7%-6.3%	4.45219e-12
Biological process - level 7 (17.93% - 8.28%)	macromolecule complex assembly (GO:0065003)	91.7%-8.3%	1.17572e-10
Biological process - level 8 (17.24% - 6.84%)	chromatin assembly or disassembly (GO:0006333)	93.75%-6.25%	1.88619e-12
Biological process - level 8 (17.24% - 6.84%)	protein-DNA complex assembly (GO:0065004)	93.92%-6.08%	1.88619e-12
Biological process - level 9 (12.41% - 2.63%)	chromatin assembly (GO:0031497)	90.36%-9.64%	1.38585e-10
5000 bp downstream (genes)			
50000 bp downstream (64 genes)			
100000 bp downstream (117 genes)			

VII. Eztabaida orokorra

Lan honetan analisi ezberdinen bidez erretravirus endogenoen (ERV) ezagutza sakontzeko hainbat alderdi aztertu dira (VII-1. irudia). Lehenik, behiaren eta zaldiaren genomatan ERVen

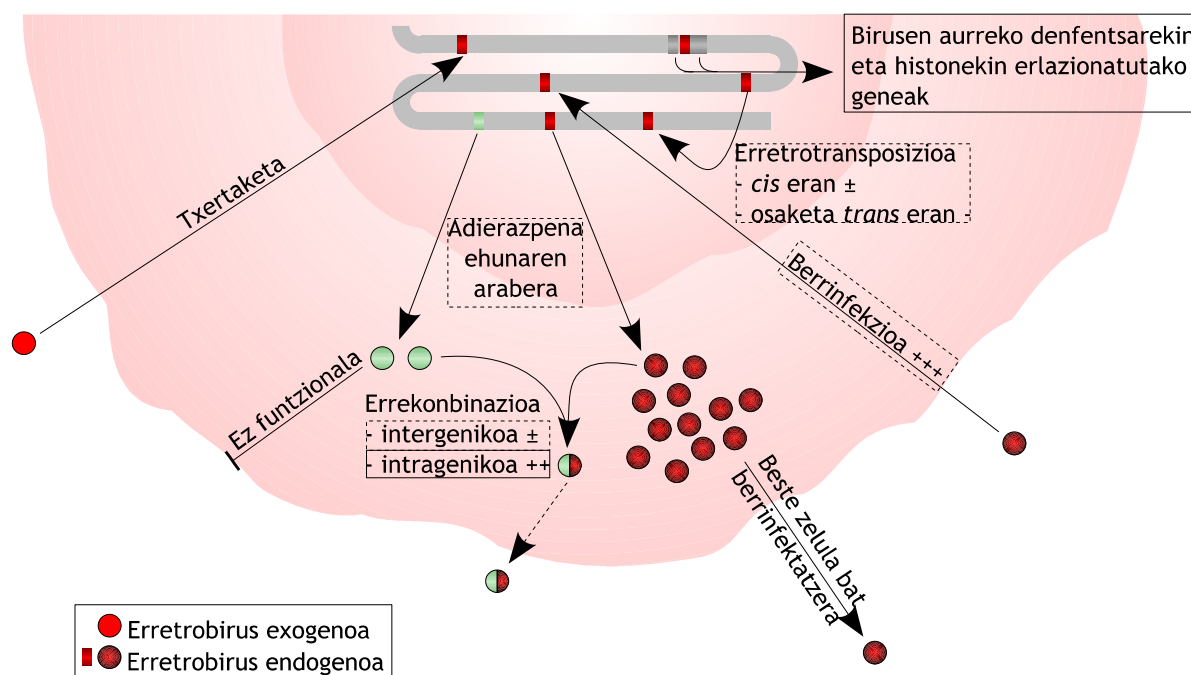


VII-1. irudia: Tesi-lan honetan egindako analisisen adierazpen eskematikoa.

● Lan honetan garatutako script-ak adierazitako pausua egiteko edota errazteko.

detekzioa hiru metodoen bidez burutu da eta analisi filogenetikoen bidez detektatutako ERVak karakterizatu dira. Bigarrenik, behi eta zaldien ERVak aztertu dira euren dinamika ebolutiboa analizatzeko. Honetarako, ERVen kopia-kopuruaren emendioa eta baimendutako sekuentzia-aldaketa aldaketak aztertu dira ERVetan errekonbinazio intergenikoa, errekonbinazio intragenikoa eta hautespena analizatu dira. Analisi hauen emaitzak konparatu ahal izateko eta joerak aurkitzeko, beste bost mamalioen (txakur, gizaki, txinpantze, sagu eta arratoi) ERVak aztertu dira. Hirugarrenik, behiaren ERVen kokapena aztertu da ERVetatik gertu zeuden geneak detektatuz eta gene hauen funtzioa aztertu da. Laugarrenik, eta azkenik, behiaren ehunetan ERVak adierazten ote ziren ikertu da behiaren EST (*expressed sequence tags*) datubasean bilaketak burutuz.

ERVak detektatzeak bere zailtasunak ditu elementu hauen kontserbazio-maila dela eta. Antzekoak diren beste lanekin alderatuz, zeintzuetan programa bakarra erabili den, lan honetan ERVak detektatzeko hiru metodo ezberdin erabili ziren. Harrigarriki, metodo ezberdinen emaitza komunak urriak izan ziren, bai aztertutako bi espezie ezberdinen genometan bai espezie bereko bi genomabertsioetan ere. Gainera, azken kasu honetan, genoma-bertsioa aldatzerakoan kromosoma batzuetan ERV gutxiago detektatu ziren. Detekzio-metodo bakoitzak irizpide ezberdinak erabiltzen dituenenez (McCarthy & McDonald, 2003; Sperber et al., 2007) ezberdintasunak egotea itxaron daiteke, baina adostasun gutxi egotea eta genomaren bertsioz bertsio galerak egotea nahiko deigarria da. Honekin lotuta dinamika ebolutiboa aztertzeko erabilitako ERVen detekzioa legoke. ERV hauek metodo bakarrarekin eta irizpide zorrotz batekin 7 genoma ezberdinetan bilatu ziren eta, itxaron daitekeen moduan, genoma bakoitzean irizpideak betetzen zituzten ERV kopurua ezberdina izan zen. Gainera, talde taxonomiko bereko bi espezieen artean (gizaki/txinpantze eta sagu/arratoi) hobeto sekuentziatuta dauden genometan ERV gehiagok bete zuen irizpidea, eta hortaz, ezin dira erabili intentsiboki sekuentziatuta ez dauden genomak. Agian sekuentziazio proiektuen barnean elementu errepikakorren kokapenean zailtasun handiagoak egoteak azal lezake gertaera hau (International Human Genome Sequencing Consortium, 2001). Hortaz, ERVen benetako kopuruaz gain, genomaren sekuentziazioaren kalitateak detekzioa baldintza lezake. Edonola ere, ERVei buruzko ezagumendua eta detekzio metodoen ezaugarriak dira detekzioaren muga nagusiak, batez ere ondo kontserbatutako ERVak detektatzen baitituzte (McCarthy & McDonald, 2003; Sperber et al., 2007).



VII-2. irudia: Erretrobirus endogenoei buruzko ikuspegi orokorra.

Atal honetan laburbildutako informazioan eta eztabaidan oinarrituta. Jarraikako lauki barnean, lan honetan lehen aldiz ERVetan deskribatzen den gertaera; etenako lauki barnean, ERVetan ezaguna zena lan honetan beste espezieetako ERVetan ikusi diren gertaerak.

Aipatutako analisiak erabilia ERVen ezagumenduan sakontzeko gai izan gara (VII-2. irudia), ERVei buruz ezaguna zena beste espezie batzuetara zabalduz, bereziki behietara, eta analisi hauek uztartuz ERV eta ostalari arteko erlazioan sakondu dezakegu.

Detektatzeko gai izan garen ERVek ez zuten banaketa genomiko homogeneorik azaldu. Gainera, ERV kopurua behian eta zaldian kromosomaren tamainarekin korrelazio positiboa zutela ikusi zen eta behian GC-edukiarekin korrelazio negatiboa. ERVak, eretrovirus exogenoen (XRV) eta berrinfekzioen arrastoak diren heinean, XRVek azaltzen dituzten lehentasunen isla izan daitezke. XRVak genomak txertatzeko orduan tropismo ezberdinak azaltzen dituzte (Bushman et al., 2005), adibidez, GIBak (gizakiaren immunoeskasiaren birusa) adierazpen aktiboko guneetan, hots, adierazten diren geneetan, txertatzeko joera dauka. Hortaz, ERVen kokapen genomikoarekin lotuta ere behien ERVetatik gertu kokatzen diren geneak egon daitezke.

ERVetatik gertu kokatzen diren geneen artean, ERV/gene asoziazio gehienak gainjartzen direnean gertatu ziren. Horietatik ERVak behiaren geneen introietan txertatuta zeudenean, ERVek transkripzio norabidearen aurka txertatzeko joera azaldu zuten, gizakien (Medstrand et al., 2002; van de Lagemaat et al., 2003), saguen (van de Lagemaat et al., 2003) eta behien (Almeida et al., 2007) elementu transposakorretan oro har ezaguna zena baina ERVetan soilik gizakietan ezaguna zena. Joera hau ERVen biziraupeanarekin erlazionatu da (van de Lagemaat et al., 2006), introietan norabide berdinean txertatzea ostalariarentzat kaltegarria izan baitaiteke. Bestalde, ERVetatik gertu kokatzen ziren geneen artean birusen aurreko defentsan eta histonekin erlazioa zutenak gainerrepresentatuta zeuden eta gene hauen parte bat auresandako geneak edo pseudogeneak izan ziren. Elementu errepikakorren bidez sortzen diren eretrogeneak eta pseudogene prozesatuek gizaki eta saguen genomak gene duplikazioek baino eragin handiagoa izan zezaketela aditzera eman da (Sakai et al., 2007) eta, hortaz, auresandako gene eta pseudogene hauek duplikazio saiakeren ondorio izan daitezke. Gainera, gizakian ERVen bidezko gene duplikazioak (Doxiadis et al., 2008) eta geneen trasdukzioak (Jamain et al., 2001) deskribatu dira eta behian balizko duplikazio gertaerak detektatu ditugunez, ERVen bidezko duplikazioak mamalioen genomak ezaugarri bat izan litezke.

Geneetatik gertu txertatzeak BoERVek adierazi izanarekin lotura izan dezake, *Arabidopsis*en elementu transposakorrek proposatu den bezala (Hollister & Gaut, 2009). Gizakietan ERVak ehun eta garapen fasearen arabera adierazten direla ezaguna zen (Okahara et al., 2004) eta lan honetan behiaren ehunetan ERVen adierazpena aldakorra zela ikusi da. Gainera, ERV familia eta gene eretrobiral bakoitzaren azterketa ere burutu zen. Dударik gabe *env* genea da gehien adierazi zena, batez ere enbrioiaren ehunetan, gizakian edota ardian ezaguna zen gertaera (Spencer et al., 2003; Okahara et al., 2004). Bestalde, *env* genea ez zen behiaren plazentan gizakian, saguan edo ardian (Jern & Coffin, 2008) beste adierazi. ERVen adierazpen honetan metilazioa gakoa dela aditzera eman da (Okahara et al., 2004; Maksakova et al., 2008) eta espezie artean metilazio patroia ezberdinak ematen direnez (Dindot et al., 2004; Curchoe et al., 2005) ezberdintasunak azal litzake. Gainera, metilazio patroia familia edo eskualde genomikoarekin baino ERVen LTR funtzional eta ehunarekin erlazionatu dira (Gimenez et al., 2009).

Gene eretrobiralen adierazpenarekin jarraituz, enbrioiaren *pol* genearen adierazpena nabarmena izan zen. Alde batetik, saguen enbrioiaren *pol* geneak kodetzen duen alderantzizko transkriptasa enbrioiaren garapen normalerako beharrezkoa dela ondorioztatu zen (Spadafora, 2004). Bestetik, *pol*en aktibitateak berrinfekzioak (Seifarth et al., 2005) eta errekonbinazioak (Telesnitsky & Goff, 1997) erraz ditzake.

Dinamika ebolutiboan sakonduz, berrinfekzioa eta errekonbinazioa nahiko ohikoak direla dirudite eta, esan bezala, *pol*en adierazpenarekin lotuta egon litezke. Oro har, ERVen kopuruaren emendioa berrinfekzioan oinarrituta legoke eta eretrotransposizioa, bai cis zein trans eran, ez luke eragin handirik edukiko kopuruaren emendioan, egindako analisisetan egon zitezkeen alborapenak ahaztu gabe. Bestalde, gene arteko errekonbinazio-gertaera gutxi detektatu ziren baina bai, ordea,

geneen barneko errekonbinazioak, azken mekanismo hau lehen aldiz modu sistematiko batean lan honetan aztertua izan delarik. Bai bata zein bestearen ondorioz, hots, errekonbinazioaren ondorioz, funtzionalak diren sekuentziekin gertatzen baldin bada, ERV ez-funtzionalek funtzionaltasuna berreskura lezaketela aditzera eman da (Gifford & Tristem, 2003). Bestalde, errekonbinazio ez homologoak genomaren berrantolaketa iturri moduan proposatu dira (Jern & Coffin, 2008). Gainera, ERVetan baimenduta dauden aldaketak aztertzeko, errekonbinazio intragenikoaz gain, hautespena erabili zen. Kodoi gehienak hautespen purifikatzailepean egon baziren ere, hautespen neutralpean eta positiboan zeudenak ere detektatu ziren, oro har, espezie eta ERV gehienetan. Hortaz, hainbat kasutan hautespenaren erlaxazioa gertatzen ari dela ondorioztatu daiteke. Beraz, ERVek funtzionaltasuna galtzeko bidean egon daitezkeela iradoki dezake. Agian, behaketa hau pseudogeneek funtzionaltasuna galtzen dutenean jasotzen duten (Torrents et al., 2003) antzeko prozesua izan daiteke.

Hala ere, ezin daiteke ERVen dinamika orokorra ulertu elementu transposakorren (TE, *transposable elements*) testuinguruan ez bada. TEak, oro har, genoma-ostalariak inbaditzeko eta moldatzeko gai dira eta genomak defentsa-mekanismoak garatu dituzte TEen aurrean defendatzeko. (Biemont & Vieira, 2006; Blikstad et al., 2008). Adibidez, metilazioa TEen adierazpena kontrolatzeko sortutako mekanismo bat izan zela proposatu zen (Yoder et al., 1997).

Orain arte aipatutako emaitzen interpretazioak uztartuz bi agertoki azaleratzen dira: ERV (edo TE) eta geneen arteko lehia edota ERV (edo TE) eta genomaren arteko kolaborazioa (VII-1. taula).

VII-1. taula: ERV eta genomaren arteko erlazioen laburpena.

	Lehia	Kolaborazioa	Lan honetan ikusitakoa
ERV kopurua	· Emendioak genoma guztiz kolonizatzea (eta suntsitzea) eragin dezake. ^a		· Aldakorra espeziaren eta familiaren arabera: behien zaldian baino gehiago. · Aztertutako espezie gehienetan emendioa berrinfekzio bidez.
ERVen errekonbinazioa	· Gaitasun patogenikoak berreskura ditzakete. ^b · Errekonbinazio ez-legitimoei genoma desantola dezakete. ^c		· Gene arteko errekonbinazioa ez oso ohikoa. · Gene barruko errekonbinazio-gertaerak nahiko ohikoak direnez ERVek beren egitura berreskura lezaketete.
ERVen kokapena	· Geneen funtzioa eten dezakete. ^{c,d}	· Geneetatik zentzuzko distantzia batera kokatuz erregulazio eskualde berriak, motzitsasketa alternatiboak, etab eman ditzakete. ^{c, d, e} · Berrantolaketa genomikoetarako gune aproposak izan daitezke. ^c · Geneen duplikazioan lagundu dezakete. ^{f, g}	· Behien introietan ERVak transkripzio kontrako norabidean txertatzen dira. · ERVetatik gertu dauden behien geneak birusen aurkako defentsarekin eta histonekin erlazionatuta daude. · ERVetatik gertu dauden behien gene proportzio bat pseudogene edo auresandako geneak dira, duplikazio saiakeren arrastoak izan daitezkeenak.
ERV adierazpena	· Ez bada erregulatzen kaltegarria euren jatorri patogenikoa dela eta. ^b	· Erregulatzen bada onuragarria: <i>pol</i> eta <i>env</i> enbrioian. ^{b, c}	· Behietan, garapen-fase eta ehunaren arabera ERV familia bakoitzaren geneak modu ezberdinean adierazten dira, nagusiki enbrioian eta guriin endokrinoetan. Hots, adierazpena kontrolatuta dago.

^a (Katzourakis et al., 2005); ^b (Gifford & Tristem, 2003); ^c (Jern & Coffin, 2008); ^d (van de Lagemaat et al., 2006); ^e (Belshaw et al., 2007); ^f (Jamain et al., 2001); ^g (Doxiadis et al., 2008)

Alde batetik, ERVen kopuruaren emendioa ez da genomentzat batere onuragarria, genoma suntsitzera hel daitezkeelako (Katzourakis et al., 2005). Beraz, alde batetik, genomak txertaketa berriak ekidin behar ditu, adibidez, errestrikzio faktore erretrobiralen bidez (Goff, 2004) eta,

bestetik, ERVen adierazpena kontrolatu (Maksakova et al., 2008). Adierazpen kontrolaren garrantzia hainbat gaixotasunetan gertatzen den ERVen gainadierazpenean edo adierazpen patroiaaren aldaketan ikus daiteke, esate baterako, minbizian (Stauffer et al., 2004; Frank et al., 2008; Golan et al., 2008). Bestalde, ez-funtzionalak diren ERVak kontrolik gabe adierazten badira, beste ERV batzuekin errekonbina dezakete eta funtzionaltasuna berreskuratu. Honen ondorioz, beste zelula bat infektatu dezake berrinfekzioaren bidez edo XRV moduan aska daiteke (Gifford & Tristem, 2003). Bestalde, txertaketa berriek geneen funtzionamendua oztopatzearen arriskua daukate ere (van de Lagemaat et al., 2006; Jern & Coffin, 2008), batez ere XRVak erakusten dituzten tropismoak ikusita (Bushman et al., 2005). Azkenik, ERVek duten antzekotasunean oinarrituta, genomaren puntu ezberdinean dauden ERV artean errekonbinazioa gerta daiteke, genomaren egitura aldatuz eta, hortaz, ostalariarengan eragin kaltegarriak sortuz. Adibidez, gizakiaren Y kromosoman gertatzen den delezio batek antzutasuna eragiten du (Blanco et al., 2000).

Bestetik, ERVen LTRek geneetan modu onuragarrian eragin dezakete erregulazio eskualde berriak, hala nola, promotore eta areagotzaileak, eman baititzake edo motzitsasketa alternatibo leku berriak (van de Lagemaat et al., 2006; Belshaw et al., 2007; Jern & Coffin, 2008). Baita ere, berrantolaketa genomikoetan parte har dezaketela aditzera eman da (Jern & Coffin, 2008) eta, horrela, genomari plastikotasuna eman. Honetaz gain, lehenago azaldu bezala, ERVak gene duplikazioetan parte har lezakete (Jamain et al., 2001; Doxiadis et al., 2008). Azkenik, ERVen adierazpen kontrolatua ikusi da eta, honek, ostalariari ematen dion onurekin lotuta egon liteke (Gifford & Tristem, 2003; Jern & Coffin, 2008).

Behien ERVei dagokionez, lan honen bidez, esperimentalki lortutako informazioa (Xiao et al., 2008) emendatzea lortu da, deskribatuta zauden 4 ERV familietatik gain (Xiao et al., 2008), beste 20 familia deskribatu baitira, horien artean behiaren genomari familia ugariena (BoERV1 bezala izendatu dena) dirudiena deskribatuz. Zaldien ERVei dagokiela, ordea, gutxi zen ezagun, eta lan honi esker, 15 ERV familia daudela jakin da. Hala ere, zaldien behian baino ERV gutxiago detektatu ziren. Bestalde, behien eta zaldien ERV gehienak euren kopurua berrinfekzio bidez emendatu zuten, txakur, gizaki, txinpantze eta sagu ERV gehienak bezala. Salbuespen bakarra arratoiaren ERVak izan ziren, lan honetan lehengoak deskribatzen den gertaera. Hala ere, arratoiaren ERVei buruzko ezagumenduak gertaera hau azaltzeko arrazoiak proposatzea ez du ahalbidetzen. Bestalde, behien eta zaldien ERVen kontserbazioari dagokionez, hautespena purifikatzailea bada ere, hautespena erlaxatua zuten kodoiak azaldu ziren, gainontzeko espezieetan bezala. Kontserbazioarekin lotuta, errekonbinazio intrageniko gertaerak detektatu ziren, behian (gizaki, txinpantze, sagu, arrato eta txakurren parekoa) zaldien baino gehiago. Behiaren kasuan, ERVak birusaren aurreko defentsarekin eta histonekin erlazionatutako geneetatik hurbil txertatuta daudela ikusi zen, lehendik ERVetan ezaguna ez zen gertaera bat. Bestalde, behiaren ERVak geneen introietan kontrako norabidean txertatzen dira, behian elementu transposakor klase batzuetan ezaguna zena eta gizaki eta saguen elementu transposakorretan (Medstrand et al., 2002; van de Lagemaat et al., 2003). Azkenik, behiaren ehunetan ERVak adierazten dira, gizakian gertatzen den bezala (Stauffer et al., 2004; Seifarth et al., 2005), baina ezberdintasun batzuk azalduz, adibidez, plazentan.

Laburbilduz, aztertu diren alderdi ezberdinetatik lortutako emaitzak uztartuz oraindik txatal gutxi batzuk ezagutzen zaizkion erretratu konplexu eta interesgarri bat azaleratzen da. Urteetan zehar DNA-zabortzat hartu diren elementu genomiko hauek ez dira soilik mutazioak pairatzen duten eta ez-aktiboak diren sekuentziak. Benetan, eretrovirus endogenoak genoma ostalarietan dinamika orokor bizi eta konplexua azaltzen duten elementu genomikoak dira eta genomarekin etengabeko erlazio konplexuak dituzte.

Gainera, lan hau genomikan kokatzen den heinean, oraindik behaketak egiten diren garaietan gaude, etorkizunerako hainbat galdera eta gai interesgarriak ateak zabalik utziz, adibidez:

- ERVak hautespen purifikatzailepean eta positiboan egotearen jatorria eta genomaren ostalariarentzat dituen inplikazioak.

- ERVen txertaketa-preferentziarekin erlazionatuta dauden mekanismoak.
- ERVen adierazpen kontrolatuarekin erlazionatuta dauden mekanismoak.

Azkenik, lan honetan ERVe buruzko eta genoma-ostalariarekin duten erlazioei buruzko ezagumenduan sakontzeko lehen pausuak eman ditugu eta hauek etorkizunean egin daitezkeen aipaturiko analisi berrien eta bestelakoen abiapuntu izan daitezke.

Bibliografia

- Almeida, L., I. Silva, W. Silva, J. Castro, P. Riggs, C. Carareto, and M. Amaral. 2007. The contribution of transposable elements to *Bos taurus* gene structure. *Gene* **390**:180-189.
- Belshaw, R., J. Watson, A. Katzourakis, A. Howe, J. Woolven-Allen, A. Burt, and M. Tristem. 2007. Rate of recombinational deletion among human endogenous retroviruses. *Journal of Virology* **81**:9437-9442.
- Biemont, C. and C. Vieira. 2006. Genetics: Junk DNA as an evolutionary force. *Nature* **443**:521-524.
- Blanco, P., M. Shlumukova, C. A. Sargent, M. A. Jobling, N. Affara, and M. E. Hurles. 2000. Divergent outcomes of intrachromosomal recombination on the human Y chromosome: male infertility and recurrent polymorphism. *J Med Genet* **37**:752-758.
- Blikstad, V., F. Benachenhou, G. Sperber, and J. Blomberg. 2008. Evolution of human endogenous retroviral sequences: a conceptual account. *Cellular and Molecular Life Sciences* **65**:3348-3365.
- Bushman, F., M. Lewinski, A. Ciuffi, S. Barr, J. Leipzig, S. Hannenhalli, and C. Hoffmann. 2005. Genome wide analysis of retroviral DNA integration. *Nature Reviews Microbiology* **3**:848-858.
- Curchoe, C., S. Zhang, Y. Bin, X. Zhang, L. Yang, D. Feng, M. O'Neill, and X. Tian. 2005. Promoter-specific expression of the imprinted IGF2 gene in cattle (*Bos taurus*). *Biology of Reproduction* **73**:1275-1281.
- Dindot, S., K. Kent, B. Evers, N. Loskutoff, J. Womack, and J. Piedrahita. 2004. Conservation of genomic imprinting at the XIST, IGF2, and GTL2 loci in the bovine. *Mammalian Genome* **15**:966-974.
- Doxiadis, G., N. de Groot, and R. Bontrop. 2008. Impact of endogenous intronic retroviruses on major histocompatibility complex class II diversity and stability. *Journal of Virology* **82**:6667-6677.
- Frank, O., C. Verbeke, N. Schwarz, J. Mayer, A. Fabarius, R. Hehlmann, C. Leib-Mosch, and W. Selfarth. 2008. Variable transcriptional activity of endogenous retroviruses in human breast cancer. *Journal of Virology* **82**:1808-1818.
- Gifford, R. and M. Tristem. 2003. The evolution, distribution and diversity of endogenous retroviruses. *Virus Genes* **26**:291-316.
- Gimenez, J., C. Montgiraud, G. Oriol, J. P. Pichon, K. Ruel, V. Tsatsaris, P. Gerbaud, J. L. Frendo, D. Evain-Brion, and F. Mallet. 2009. Comparative Methylation of ERVWE1/Syncytin-1 and Other Human Endogenous Retrovirus LTRs in Placenta Tissues. *DNA Res* **16**:11-21.
- Goff, S. 2004. Retrovirus restriction factors. *Molecular Cell* **16**:849-859.
- Golan, M., A. Hizi, J. Resau, N. Yaal-Hahoshen, H. Reichman, I. Keydar, and I. Tsarfaty. 2008. Human endogenous retrovirus (HERV-K) reverse transcriptase as a breast cancer prognostic marker. *Neoplasia* **10**:521-5U3.
- Hollister, J. D., B. S. Gaut. 2009. Epigenetic silencing of transposable elements: a trade-off between reduced transposition and deleterious effects on neighboring gene expression. *Genome Res* **19**: 1419-1428.
- International Human Genome Sequencing Consortium. 2001. Initial sequencing and analysis of the human genome. *Nature* **409**:860-921.
- Jamain, S., M. Girondot, P. Leroy, M. Clergue, H. Quach, M. Fellous, and T. Bourgeron. 2001. Transduction of the human gene FAM8A1 by endogenous retrovirus during primate evolution. *Genomics* **78**:38-45.
- Jern, P. and J. Coffin. 2008. Effects of Retroviruses on Host Genome Function. *Annual Review of Genetics* **42**:709-732.
- Katzourakis, A., A. Rambaut, and O. Pybus. 2005. The evolutionary dynamics of endogenous retroviruses. *Trends in Microbiology* **13**:463-468.
- Maksakova, I., D. Mager, and D. Reiss. 2008. Keeping active endogenous retroviral-like elements in check: the epigenetic perspective. *Cellular and Molecular Life Sciences* **65**:3329-3347.
- McCarthy, E. and J. McDonald. 2003. LTR_STRUC: a novel search and identification program for LTR retrotransposons. *Bioinformatics* **19**:362-367.
- Medstrand, P., L. van de Lagemaat, and D. Mager. 2002. Retroelement distributions in the human genome: Variations associated with age and proximity to genes. *Genome Research* **12**:1483-1495.
- Okahara, G., S. Matsubara, T. Oda, J. Sugimoto, Y. Jinno, and F. Kanaya. 2004. Expression analyses of human endogenous retroviruses (HERVs): tissue-specific and developmental stage-dependent expression of HERVs. *Genomics* **84**:982-990.
- Sakai, H., K. Koyanagi, T. Imanishi, T. Itoh, and T. Gojobori. 2007. Frequent emergence and functional resurrection of processed pseudogenes in the human and mouse genomes. *Gene* **389**:196-203.

- Seifarth, W., O. Frank, U. Zeilfelder, B. Spiess, A. Greenwood, R. Hehlmann, and C. Leib-Mosch.** 2005. Comprehensive analysis of human endogenous retrovirus transcriptional activity in human tissues with a retrovirus-specific microarray. *Journal of Virology* **79**:341-352.
- Spadafora, C.** 2004. Endogenous reverse transcriptase: a mediator of cell proliferation and differentiation. *Cytogenetic and Genome Research* **105**:346-350.
- Spencer, T., M. Mura, C. Gray, P. Griebel, and M. Palmarini.** 2003. Receptor usage and fetal expression of ovine endogenous betaretroviruses: Implications for coevolution of endogenous and exogenous retroviruses. *Journal of Virology* **77**:749-753.
- Sperber, G., T. Airola, P. Jern, and J. Blomberg.** 2007. Automated recognition of retroviral sequences in genomic data - RetroTector (c). *Nucleic Acids Research* **35**:4964-4976.
- Stauffer, Y., G. Theiler, P. Sperisen, Y. Lebedev, and C. V. Jongeneel.** 2004. Digital expression profiles of human endogenous retroviral families in normal and cancerous tissues. *Cancer Immun* **4**:2.
- Telesnitsky, A. and S. Goff.** 1997. Reverse transcriptase and the generation of retroviral DNA, p. 121-160. *In* J. Coffin, S. Hughes, and H. Varmos (eds.), *Retroviruses*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Torrents, D., M. Suyama, E. Zdobnov, and P. Bork.** 2003. A Genome-Wide Survey of Human Pseudogenes. *Genome Research* **13**:2559-2567.
- van de Lagemaat, L., J. Landry, D. Mager, and P. Medstrand.** 2003. Transposable elements in mammals promote regulatory variation and diversification of genes with specialized functions. *Trends in Genetics* **19**:530-536.
- van de Lagemaat, L., P. Medstrand, and D. Mager.** 2006. Multiple effects govern endogenous retrovirus survival patterns in human gene introns. *Genome Biology* **7**.
- Xiao, R., K. Park, H. Lee, J. Kim, and C. Park.** 2008. Identification and classification of endogenous retroviruses in cattle. *Journal of Virology* **82**:582-587.
- Yoder, J., C. Walsh, and T. Bestor.** 1997. Cytosine methylation and the ecology of intragenomic parasites. *Trends in Genetics* **13**:335-340.

VIII. Ondorioak

Lan honetan egindako analisi eta interpretazioetan oinarrituta, ondorio hauek lortu ditugu:

1) Erretrobirus endogenoak detektatzeko orduan metodo ezberdinak erabiltzea gomendagarria da, metodo bakoitzak elementu ezberdinek detektatzen baititu. Gainera, metodo bakoitzaren mugez gain, genomaren sekuentziazio-mailak erretrobirus endogenoen detekzioa baldintzatzen du, genoma-bertsio batetik bestera metodoaren arabera detektatutako ERVen kopuruak gutxienez %23ko emendioa izan baitu.

2) Detektatu diren behiaren erretrobirus endogenoak I eta II klaseetan sailkatzen diren 24 familietan bana daitezke, hauetatik 20 orain arte ezezagunak. Horien artean BoERV1 bezala izendatu den familia dago, behiaren genomaren ugariena dena eta Ruminantia-ren espezifikoa izan daitekeena. Ez da III klaseko elementurik detektatu.

3) Detektatu diren zaldiaren erretrobirus endogenoak I, II eta III klaseetan sailkatzen diren 15 familietan bana daitezke. Lan hau zaldiaren ERVen familiak deskribatzen lehena da, lehendik esperimentalki lortutako sekuentzien eta gainontzeko ERVen arteko erlazio filogenetikoa finkatuz.

4) Aztertutako espezie bien erretrobirus endogenoak konparatzerakoan, erretrobirus endogenoen kopurua, familia-kopurua eta sailkatzen diren klaseak ezberdinak izan ziren, erretrobirus endogenoek genoma bakoitzean presentzia ezberdina azalduz. Ezberdintasun hauek espezie bakoitzaren bilakaeran eta genomaren dinamikaren menpekoak izan daitezke.

5) Behiaren eta zaldiaren erretrobirus endogeno gehienek, txinpantze, sagu eta txakurretan bezala, eta gizakietan ezaguna zen bezala, euren kopurua berrinfekzio bidez emendatu dute, hiru gene erretrobiral nagusiak hautespen purifikatzailepean baitzeuden. Salbuespen bakarra arratoiaren erretrobirus endogenoak dira, gure analisisen arabera batez ere transposizio bidez emendatu direnak.

6) Hiru gene erretrobiral nagusien sekuentzian zeharreko errekonbinazioa intragenikoaren analisiek, behiaren erretrobirus endogenoen geneen hainbat eskualdek, gizaki, txinpantze, sagu, arratoi eta txakurrak bezala, errekonbinazio-gertaerak jasan dituztela jarri dute agerian. Zaldien ERVetan, aldiz, askoz gertaera gutxiago detektatu ziren.

7) Hiru gene erretrobiral nagusien sekuentzian zeharreko hautespenaren analisiek, behiaren eta zaldiaren erretrobirus endogeno gehienek geneak, gizaki, txinpantze, sagu, arratoi eta txakurretan bezala, hautespen purifikatzailepean daudela adierazten dute. Hala ere, hautespen neutralpean dauden kodoiak ere detektatu dira, hautespenaren erlaxazioa iradoki dezakeenak. Baita ere hautespen positiboan dauden kodoiak detektatu dira, gene baten kodoien %30a izatera hel daitekeena. Honek funtzio berriak lortzea ahalbidetuta egon litekeela iradoki lezake.

8) Behiaren erretrobirus endogenoak batez ere behiaren hainbat geneen introietan kokatuta daude, eta kopuru txikiago batean geneetatik gertu edo geneak inguratzen. Geneen introietan kokatzen direnean, transkripzioaren aurkako norabidean txertatzeko joera azaltzen dute nagusiki, norabide berean txertatzeak genearen funtzioa eten dezakeelako eta, ondorioz, hautespen negatiboan egon daitezke.

9) Babelomics programaren gainerrepresentatutako terminoen analisisan oinarrituta, behiaren erretrobirus endogenoetatik gertu agertzen diren geneak birusen aurreko defentsarekin eta histonekin erlazioa dute, beste batzuen artean. Erretrobirusen txertaketa garaian gene hauek aktibo egoteak edota erretrobirusek ezagutu dezaketen ezaugarriak izateak gertaera hau azal lezakete.

10) Behiaren erretrobirus endogenoetatik gertu agertzen diren hainbat gene pseudogene (%12 eta %30 artean) edo auresandako geneak dira. Hauek gene-duplikazioen gertaeren arrastoak izan daitezke.

11) Lan honetan, lehen aldiz, behiaren ehunetan behiaren erretrobirus endogenoak adierazi egiten direla deskribatzen da. Adierazpena garapen-fase, ehun eta gene erretrobiralaren arabera aldatzen da; hortaz, behiaren ehunetan behiaren erretrobirus endogenoen adierazpena kontrolatuta dago. Gainera, erretrobirus endogenoen adierazpena, bereziki *env* genearena, batez ere guruin

Ondorioak

endokrinoetan eta enbrioian gertatzen da, enbrioiaaren defentsarekin lotuta egon daitekeena.

12) Lan honetan erretrobirus endogenoen hautespen maila, txertaketa preferentziak, eta adierazpen kontrolatua bezalako gertaerak ikusi dira. Behaketa hauek erretrobirus endogeno eta genoma-ostalariaren arteko erlazioa kooperazio eta lehia arteko orekan egotea iradoki lezake eta, hortaz, erlazio hau dinamika bizi eta konplexu bat deneko ideia indartzen dute.

