XANTARES BIOCIENTÍFICOS

Dende o grupo de traballo de formación e divulgación de **BioReDes** queremos anunciar a terceira charla encadrada nos ***Xantares Biocientíficos***, que terá lugar o **venres 7 de xuño ás 13:00 horas no edificio CACTUS**. A intervención levará por título **"La contribución de la genética al declive y la recuperación del lince ibérico**" e será impartida polo **Dr. José Antonio Godoy**, Investigador Científico da Estación Biolóxica de Doñana (CSIC). Para que poidades ter información de antemán sobre a importancia do tema, achegámosvos catro artigos científicos e un breve resumo da presentación.

Aproveitamos para lembrar que, se tedes suxestións sobre charlas de interese para os Xantares BioCientíficos, podedes remitilas a Iago Rodríguez, técnico de I+D da Agrupación (iago.palmeiro@usc.es)

Esperámoste! Ven a tomar un petisco con BioReDes!

<u>Breve resumo da presentación</u>

"La contribución de la genética al declive y la recuperación del lince ibérico"

José Antonio Godoy

Investigador Científico de la Estación Biológica de Doñana (CSIC)

Nuestro lince ibérico ha sido durante años un dramático ejemplo de especie al borde de la extinción, habiéndose convertido en un símbolo y emblema de la conservación de especies en España y Europa. Hoy la especie va camino de transformase en un ejemplo de recuperación con escasos precedentes en el mundo. Desde hace años venimos

estudiando la variación genética en el lince ibérico con técnicas moleculares, incorporando recientemente las novedosas y poderosas aproximaciones genómicas. Con estas herramientas hemos evaluado hasta qué punto el declive sufrido por la especie ha empobrecido y deteriorado su acervo genético, y si esto ha contribuido a su casi-extinción. Los datos son concluyentes al mostrar que la diversidad de la especie es una de las más bajas reportadas hasta la fecha, como consecuencia de bajos tamaños poblacionales y de un declive continuado que se inició hace siglos. Los resultados también sugieren que la consanguinidad ha estado limitando la afectado a la supervivencia y reproducción y por tanto la capacidad de recuperación, al menos en la población de Doñana. Estas conclusiones han propiciado una gestión dirigida a favorecer el intercambio genético con la otra población remanente en Andújar, lo que parece haber mejorado la demografía y contribuido a la recuperación de esta población. Además, la información genética se está utilizando para la gestión de una población *ex situ* que está aportando individuos genéticamente sanos con los que se está reestableciendo la especie en áreas de las que llevaba décadas desaparecida. El futuro de la especie depende de que estas poblaciones crezcan rápido y se conecten genéticamente entre sí para evitar que el deterioro genético vuelva a comprometer su supervivencia. La genética, que parece haber sido parte del problema de lince ibérico, está siendo ahora un elemento clave en su recuperación.

Bibliografía:

Marmesat E, Lucena-Pérez M, Godoy JA (2018) Aplicación de la genética a la conservación del lince ibérico. Investigación y Ciencia, 32-39.

Godoy JA, Marmesat E, Lucena-Pérez M, Kleinman-Ruiz D (2017) Biología molecular en la conservación del lince ibérico. Dossier: Revolución Molecular en las Ciencias Naturales. Revista de la Sociedad Española de Bioquímica y Biología Molecular 194, 19-23.

Casas-Marcé M, Marmesat E, Soriano L, et al. (2017) Spatio-temporal dynamics of genetic variation in the Iberian lynx along its path to extinction reconstructed with ancient DNA. Molecular Biology and Evolution 34, 2893–2907.
Abascal F, Corvelo A, Cruz F, et al. (2016) Extreme genomic erosion after recurrent demographic bottlenecks in the highly endangered Iberian lynx. Genome Biology 17, 251.

CONSERVACIÓN

APLICACIÓN DE LA GENÉTICA A LA CONSERVACIÓN DEL LINCE IBÉRICO

Las técnicas genéticas, y desde hace poco las genómicas, están ayudando a diseñar estrategias para mejorar la viabilidad de la especie, que sufre un notable deterioro genético

Elena Marmesat, María Lucena-Pérez y José A. Godoy

Como consecuencia de las actividades humanas, la Tierra se halla inmersa en una dinámica de alteraciones drásticas que afectan al ambiente y al funcionamiento de sus ecosistemas, un fenómeno que es conocido como cambio global. Un componente importante de ese cambio es la pérdida de biodiversidad a causa de la extinción de especies. Ha aumentado tanto en los últimos decenios que numerosos expertos afirman que nuestro planeta está sufriendo el sexto período de extinción masiva de la historia.

Pero las especies no solo están desapareciendo. Algunas están perdiendo o reduciendo notablemente sus poblaciones, formadas por individuos que comparten un mismo espacio y ciertas características genéticas. Las poblaciones que se contraen experimentan una serie de problemas genéticos, como la pérdida de diversidad genética y la consanguinidad, que pueden hacer precipitar la extinción de la especie en poco tiempo. De ahí que en cualquier programa de conservación de una especie amenazada debería contemplarse y evaluarse su estado genético como parte integral del plan. Los estudios genéticos y genómicos pueden ayudar en esta empresa porque ofrecen herramientas para calibrar los riesgos a los que está sometida la especie y permiten diseñar estrategias de gestión que los minimicen.

El lince ibérico (*Lynx pardinus*) es uno de los felinos más amenazados del mundo y representa un símbolo de la conservación de la fauna ibérica. Aunque en el pasado se hallaba extendido por la mayor parte de la península, durante el siglo xx sufrió un pronunciado declive, principalmente a causa de la persecución directa, los cambios en el uso del suelo, que redujeron la cantidad de hábitat favorable, y el declive del conejo, su presa principal.

En su momento más crítico, a principios de los 2000, la especie quedó restringida a dos únicas poblaciones situadas en dos pequeñas áreas aisladas entre sí al sur de la península ibérica: Doñana y Sierra Morena. En ese momento, los individuos no sumaban más de cien en total. Esta situación motivó su catalogación como especie en peligro crítico de extinción por la Unión Internacional para la Conservación de la Naturaleza (UICN). Ello dio el impulso definitivo a la adopción de medidas de conservación lideradas por cuatro proyectos Life consecutivos, desarrollados entre 1994 y 2016. Los últimos tres proyectos han sido liderados por la Consejería de Medio Ambiente de la Junta de Andalucía, y el último ha contado con la participación de Portugal y cuatro comunidades autónomas (Extremadura, Castilla-La Mancha, Región de Murcia y Andalucía).

Gracias a estas actuaciones, que inicialmente se centraron en reducir la mortalidad no natural del lince y aumentar la

ANTONIO RIVAS. PROGRAMA DE CONSERVACIÓN EX-SITU DEL LINCE IBÉRICO

SALIEGA, una hembra procedente de la población de linces de Doñana, dio a luz en 2005 a la primera camada nacida en un centro de cría en cautividad. Estas y otras actuaciones están contribuyendo a la conservación de la especie.

disponibilidad de conejo, se evitó primero la extinción de las dos poblaciones remanentes y después se consiguió invertir las tendencias demográficas negativas. Más recientemente, se iniciaron reintroducciones de individuos en algunas de las zonas ocupadas en el pasado por la especie. Todo ello llevó a que la UICN reclasificara el lince en una categoría de menor amenaza en 2015, en concreto, en la de especie «en peligro». No obstante, pese a estos resultados esperanzadores, su viabilidad a medio y largo plazo sigue dependiendo en gran medida de una gestión activa y de un seguimiento continuado de sus poblaciones.

En este contexto, nuestro grupo de investigación vio en el lince ibérico un buen modelo en el que estudiar cómo el declive de la especie se refleja en sus características genéticas actuales y hasta qué punto estas pueden llevarla a la extinción. Además, nos hemos esforzado en aportar herramientas científicas que mejoren la evaluación y la gestión de la especie, con el fin de minimizar los riesgos genéticos que pueda sufrir en el futuro.

Contamos con la ventaja de que el lince ha recibido la atención de muchos otros científicos en las últimas décadas, lo que nos ha dejado abundante y valiosa información sobre su demografía, biología, ecología y comportamiento, entre otros aspectos. Además, la gestión y el seguimiento recientes a los que ha estado sometido el felino suponen a la vez una incesante fuente de datos y muestras, y una oportunidad para orientar esta gestión desde el conocimiento científico.

En nuestro trabajo nos planteamos las siguientes preguntas: ¿cuál ha sido el estado genético de las poblaciones de lince ibérico antes y después de su declive a principios de los 2000?, ¿están afectadas la reproducción y la supervivencia?, ¿qué podemos hacer desde el área de la genética para mejorar la viabilidad de la especie a medio y largo plazo?

Elena Marmesat está finalizando su tesis doctoral en la Estación Biológica de Doñana (EBD-CSIC) sobre la variación en los genes de la respuesta inmunitaria en el lince ibérico.

María Lucena-Pérez es doctoranda en la EBD-CSIC, donde estudia las variaciones en el genoma del lince ibérico y el papel de la selección natural en el mantenimiento de la diversidad genética.

José A. Godoy es investigador de la EBD-CSIC, donde lidera un grupo centrado en la aplicación de la genética y la genómica de poblaciones para entender la evolución de especies amenazadas y contribuir a su conservación. Ha coordinado el proyecto de secuenciación del genoma del lince ibérico.

ESTUDIOS GENÉTICOS Y GENÓMICOS

Con el fin de contestar a estas preguntas, intentamos primero reconstruir la demografía y la genética de las poblaciones del lince en el pasado y su evolución con el tiempo. Para ello utilizamos un amplio conjunto de datos genéticos obtenidos a partir del análisis del ADN contenido en muestras biológicas, como pelos, sangre o excrementos del animal. Las muestras no solo pertenecían a linces actuales, sino también a individuos que vivieron en el pasado. Se estudiaron ejemplares que se conservan hoy en museos y colecciones privadas, correspondientes a los últimos siglos, así como restos descubiertos en excavaciones arqueológicas o paleontológicas, datados en varios milenios. Todos los datos juntos cubren la distribución presente y pasada de la especie en la península ibérica.

Los estudios genéticos consistieron en el análisis de 36 marcadores microsatélites (STR, de *short tandem repeats*). Estos corresponden a pequeños elementos repetitivos de ADN que contienen de uno a seis pares de bases. Se trata de elementos hipervariables, no funcionales (no codifican proteínas) y se distribuyen al azar en el genoma. Se utilizan para identificar individuos y asignar sus progenitores, y dan una idea, aunque algo imprecisa, del parentesco entre individuos en una población. Además, permiten cuantificar la diversidad genética actual y, en algunos casos, compararla con la de poblaciones antiguas.

Más recientemente, hemos recurrido a los análisis genómicos, que nos han permitido obtener datos más detallados de las características genéticas de la especie. Tales estudios fueron posibles gracias a la secuenciación del genoma del lince ibérico en 2016, un proyecto coordinado por nuestro grupo y en el que participaron, entre otros, el Centro Nacional de Análisis Genómico, el Centro de Regulación Genómica, el Instituto de Biología Evolutiva (del CSIC y la Universidad Pompeu Fabra) y el Centro de Investigaciones Biológicas (CSIC).

Con la secuenciación del genoma de 11 individuos identificamos más de 1,5 millones de posiciones variables, conocidas como SNP (de *single nucleotide polimorphisms*), esto es, sitios de la secuencia donde cambia un solo nucleótido. Posteriormente seleccionamos 1500 de estas posiciones de manera que fueran especialmente variables y estuvieran bien repartidas a lo largo del genoma. Con ellas caracterizamos una muestra de más de 300 individuos, representativa de la genética actual del lince ibérico. Con esta información hemos seleccionado subconjuntos de variantes que tienen un poder muy alto para la identificación individual y la asignación de paternidades, además de ofrecer estimaciones más precisas del grado de parentesco y de la ancestría poblacional (la proporción del genoma que procede de una de las poblaciones remanentes). Estos nuevos marcadores mejorarán, por lo tanto, la gestión genética en marcha y ofrecen una alternativa potente y económica al seguimiento de las poblaciones silvestres.

DETERIORO GENÉTICO DE LAS ESPECIES AMENAZADAS

Cuando las poblaciones de una especie reducen su tamaño, comienza a cobrar protagonismo la deriva genética, un proceso

EN SÍNTESIS

Durante el siglo xx, el lince sufrió un pronunciado declive demográfico que lo llevó al borde de la extinción. La adopción de medidas de conservación a principios de los años 2000 ha paliado en parte esa grave situación.

La gestión genética está contribuyendo a la recuperación de la especie. Esta se concibió inicialmente a partir del análisis de varios marcadores microsatélites en el ADN del lince, que indicaban una muy baja diversidad y una alta consanguinidad de las dos poblaciones remanentes en Doñana y Sierra Morena.

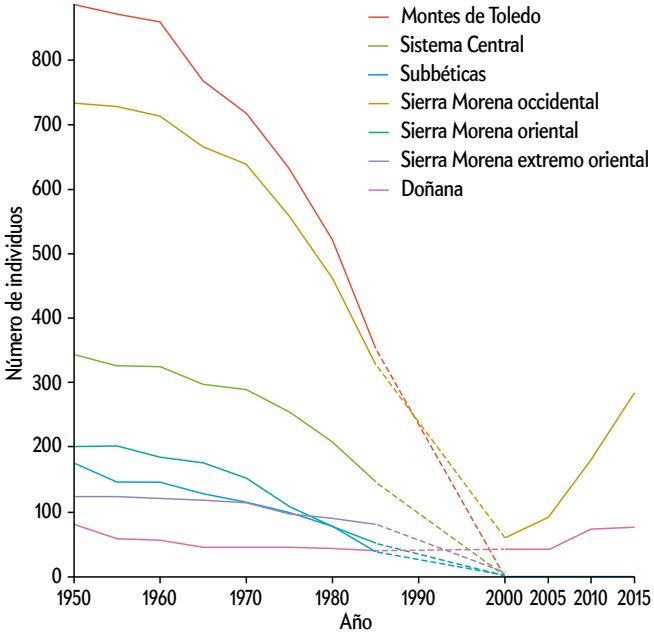
La reciente secuenciación del genoma del lince ibérico ha arrojado luz sobre la historia demográfica y evolutiva de la especie y ha permitido identificar marcadores genéticos más eficientes e informativos para el seguimiento y la gestión de las poblaciones actuales.

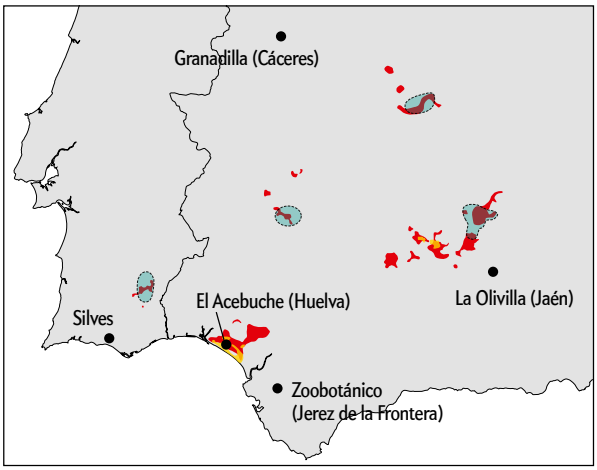
El declive del lince y su incipiente recuperación

La distribución del lince ibérico se redujo y se fragmentó enormemente en el siglo pasado. Ello provocó un grave deterioro genético de la especie, que estuvo a punto de desaparecer. Las medidas de conservación aplicadas desde inicios de los 2000, en las que se tienen en cuenta las características genéticas de las poblaciones, están logrando una paulatina recuperación.

Distribución antigua · Distribución en 1980 · Distribución en 2002

Antes del siglo XX, el lince se distribuía por toda la región mediterránea de la península ibérica, llegando a ocupar el sudeste de Francia. Después sufrió un rápido declive, y en 1980 su distribución estaba limitada al cuadrante sudoccidental, fragmentada en ocho poblaciones aisladas entre sí que sumaban menos de 2000 individuos. En 2002 solo quedaban unos 100 individuos repartidos en dos poblaciones: Doñana y Sierra Morena oriental.





Gracias a los esfuerzos de conservación, se han recuperado en parte las dos poblaciones remanentes y se han vuelto a crear otras en zonas donde la especie había desaparecido (reintroducción). Se ha logrado así que en 2017 la cifra ascendiera a casi 550 individuos en total.

Las medidas actuales consisten en la cría en cautividad en cinco centros, que producen ejemplares destinados a la reintroducción. La gestión genética permite aumentar la diversidad y minimizar la consanguinidad de las poblaciones. En ella se contempla el diseño de esquemas de apareamiento óptimo en los centros de cría, así como la selección de individuos para la reproducción en cautividad, el traslado entre poblaciones y las liberaciones en los programas de reintroducción.

Distribución en 2002 · Distribución en 2016 · Reintroducción · Centros de cría

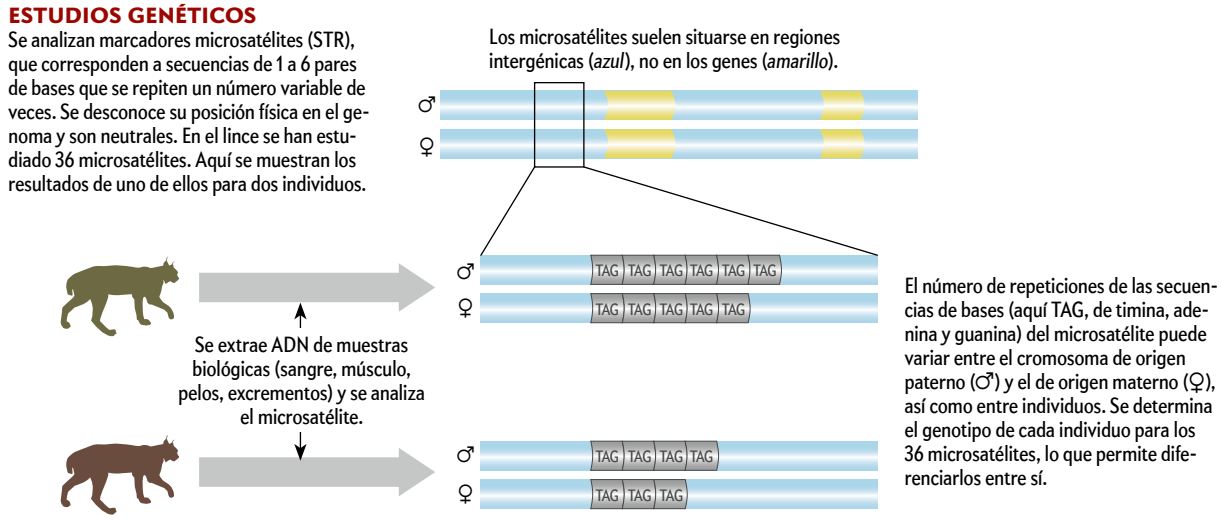
INVESTIGACIÓN Y CIENCIA, SEGÚN LOS AUTORES; DATOS DEL PROYECTO LIFE IBERLINCE (*mapa inferior*)

evolutivo impulsado por el azar que en las poblaciones grandes apenas opera.

A corto plazo, se producen dos fenómenos principales. Uno de ellos es el aumento de las variantes genéticas nocivas. Ello se debe a que en las poblaciones pequeñas la selección natural resulta menos eficaz y empiezan a acumularse variantes genéticas perjudiciales que, en condiciones normales, serían eliminadas o se mantendrían con una baja frecuencia. El segundo hace referencia a la endogamia. Cuanto más pequeña es una población, más probable es que se apareen entre sí individuos emparentados. Ello genera individuos cada vez menos diversos genéticamente, es decir, con un mayor número de genes cuyas copias (o alelos) heredadas de sus dos progenitores son idénticas. Si estas copias son además defectuosas, la función que desempeña el gen puede perderse y provocar enfermedades o malformaciones genéticas. Esta situación, conocida como depresión por endogamia o consanguinidad, actúa sinérgicamente con la acumulación de variantes perjudiciales y tiene como resultado una disminución de la eficacia biológica media de los individuos (esto es, su capacidad para

De la genética a la genómica

Para conocer la diversidad genética del lince ibérico se han llevado a cabo dos tipos de estudios: genéticos y genómicos. Los primeros han utilizado marcadores tradicionales, como los microsatélites, para diferenciar individuos de una población a partir de restos hallados en la naturaleza y para cuantificar de manera grosera la diversidad genética poblacional. Los genómicos, más novedosos, suponen un enorme salto cualitativo respecto a los genéticos, puesto que permiten conocer con mucho más detallle la variación genética funcional y valorar su efecto en la especie, además de aportar marcadores más eficientes y menos costosos. Ambos tipos de estudios pueden aplicarse a muestras antiguas para caracterizar la variación genética en el pasado y revelar los cambios que se han producido con el tiempo.



sobrevivir y reproducirse) y de la viabilidad de la población a corto o medio plazo.

Por otro lado, las reducciones más drásticas del tamaño poblacional dan lugar a cuellos de botella, en los que se produce una pérdida importante de la diversidad genética en un tiempo relativamente corto, con importantes consecuencias a largo plazo. En estas circunstancias, aunque la especie se hallara bien adaptada a su entorno actual, su baja diversidad genética no le permitiría ajustarse a los cambios ambientales futuros porque carecería de las variantes que podrían ayudarla a sobrevivir en el nuevo entorno. El riesgo de que se produzca esta situación es elevado para numerosas especies, sobre todo si tenemos en cuenta el contexto de cambio global al que se están enfrentando en nuestros días.

Estos procesos pueden continuar afectando de manera negativa a la dinámica poblacional de la especie, aunque llegaran a desaparecer los factores que originalmente la llevaron a estar amenazada. Es, por tanto, muy importante que los planes de recuperación de especies evalúen su estado genético actual y, en caso necesario, implementen medidas que ayuden a paliar estos problemas.

EL ESTADO GENÉTICO DEL LINCE

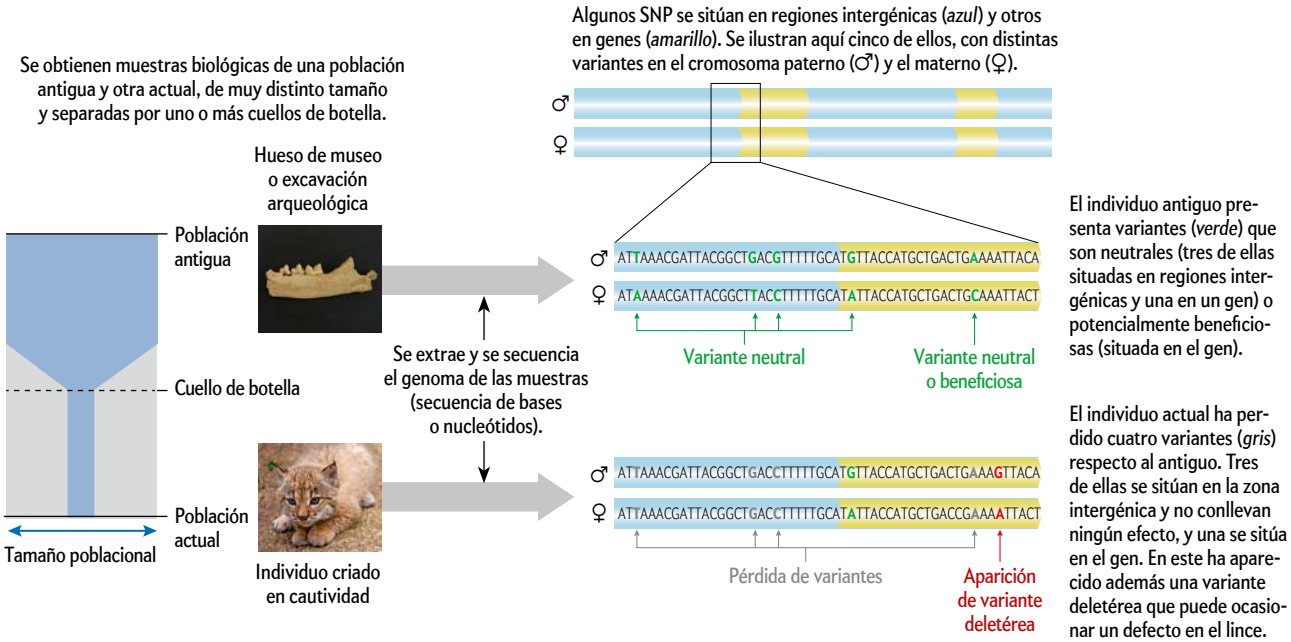
El lince ibérico posee en la actualidad una de las diversidades genéticas más bajas jamás registradas, inferior a la de otras especies amenazadas, como el demonio de Tasmania, el guepardo, el delfín del río Yangtzé o el zorro de las islas del Canal. Medimos la diversidad como la densidad media de sitios variables (SNP) respecto al total de sitios en el genoma (el total de nucleótidos). Hemos observado que en el lince ibérico esa cifra corresponde a unos cien por cada millón de sitios, mientras que en otros mamíferos, como el panda gigante y el gorila occidental, suele ser superior a mil. Especialmente acusada ha sido la pérdida de diversidad que ha sufrido la población de Doñana. Pero ¿cómo se ha llegado a esta situación? ¿Ha sobrevenido como consecuencia del declive que ha sufrido en los últimos tiempos la especie, o esta siempre ha presentado una baja diversidad?

A partir de los datos de marcadores STR y de secuencias del genoma mitocondrial completo obtenidos de las muestras fósiles y de museo, hemos explorado las dinámicas demográficas y genéticas del lince ibérico en los últimos siglos y hemos deducido que sus distintas poblaciones siguieron caminos diversos hacia su extinción. Mientras que las poblaciones periféricas de la península ibérica (Doñana, cordilleras subbéticas y Sistema Central en España, y valle del Sado en Portugal) fueron contrayéndose de forma progresiva durante la segunda mitad del siglo xx, las centrales (Montes de Toledo, Sierra Morena) permanecieron más grandes y mejor conectadas con las demás. Como consecuencia, las poblaciones periféricas se vieron más afectadas genéticamente, perdiendo diversidad y diferenciándose de otras por deriva genética, tanto más cuanto más tiempo permanecieron pequeñas y aisladas. En cambio, la población más grande y central (Montes de Toledo) no llegó a mostrar signos de erosión o deterioro genético (esto es, no disminuyó su diversidad genética ni sufrió deriva genética), probablemente porque se mantuvo grande y conectada casi todo el tiempo, para después desaparecer muy rápidamente.

Este conocimiento histórico nos ayuda a entender las causas de las diferencias genéticas actuales entre las poblaciones de Doñana y Sierra Morena. Las de Montes de Toledo y Sierra Morena

ESTUDIOS GENÓMICOS

Se analizan las secuencias de buena parte del genoma o una gran cantidad de marcadores, habitualmente SNP. Estos corresponden a nucleótidos individuales polimórficos (que presentan distintas variantes). Gracias a la disponibilidad de un genoma de referencia para la especie, podemos conocer si cambian la secuencia de una proteína y si estos cambios son deletéreos o no. Del millón y medio de SNP hallados en el lince, se han escogido unos 1500 que son muy informativos.



estuvieron fuertemente conectadas entre sí durante milenios, lo que hace que la última perdiese menos diversidad y sufriera menor deriva genética. Por el contrario, Doñana, que ha sido una población pequeña y aislada desde hace siglos, ha llegado a ser una de las poblaciones más afectadas genéticamente. La persistencia de la población de Doñana, a pesar de su extremado deterioro genético, es sin duda consecuencia de la protección conferida por el Parque Nacional y por los esfuerzos de conservación mantenidos durante décadas. Por otro lado, la extinción de la población de Montes de Toledo sin que la genética se hubiera visto afectada apunta a una rápida contracción por la presión de los factores externos ya conocidos, en este caso la destrucción de hábitat, la persecución directa y el declive de la presa.

Un resultado llamativo de la comparación de la secuencia del genoma mitocondrial completo de individuos de distintas épocas es que a principios del siglo pasado la diversidad de la especie ya era moderadamente baja. La reconstrucción de la historia demográfica da una pista del porqué: el lince ibérico no ha sido muy abundante en ningún momento y ha pasado desde su origen (que se sitúa en hace unos 300.000 años, cuando divergió del lince boreal) por sucesivos cuellos de botella que han reducido aproximadamente a la décima parte su tamaño de población. En particular, el cuello de botella documentado durante el siglo xx fue precedido por otro de similar magnitud hace unos 300 años.

Con el análisis de genomas completos han quedado al descubierto otros componentes importantes de la erosión genética. Hemos observado así que las zonas del genoma que codifican las proteínas presentan en el lince ibérico un exceso de variantes que cambian la secuencia de las proteínas y que, por lo tanto, son muy probablemente deletéreas. Estas variantes deletéreas se habrían acumulado como resultado de la menor eficacia de la selección natural «purificadora» en poblaciones pequeñas. Los análisis bioinformáticos en marcha permitirán predecir la severidad del efecto sobre la función de los genes e, idealmente, inferir las consecuencias para el fenotipo.

PÉRDIDA DE EFICACIA BIOLÓGICA

Parece claro que el lince ha sufrido erosión genética a lo largo de su historia. Lo que debemos preguntarnos ahora es si esa erosión está teniendo consecuencias en la reproducción o la supervivencia de sus individuos, es decir, si ha disminuido su eficacia biológica. Una forma habitual de evaluarlo consiste en comparar la capacidad reproductora o la supervivencia de individuos con muy distintos grados de consanguinidad. Analizar estos datos en especies amenazadas suele revestir especial dificultad, ya que todos los individuos actuales presentan cierto grado de parentesco, por lo que la variación de la consanguinidad en ellos es reducida. Además, se necesita una muestra de gran tamaño para hacer este tipo de inferencias, lo cual también resulta complicado en especies amenazadas. Pese a todo, hemos podido demostrar una correlación significativa directa entre homocigosidad (porcentaje de marcadores homocigotos en un individuo determinado; representa una medida de la consanguinidad individual) y la calidad seminal: los linces más consanguíneos tienden a tener una peor calidad del semen, lo que probablemente afecta a su capacidad reproductora.

Aunque no hayamos podido demostrarlo de forma directa, pensamos que los efectos perniciosos de la erosión genética podrían estar detrás de ciertas anomalías observadas. Por ejemplo,

CORTESÍA DE LOS AUTORES (*hueso mandíbula*); PROGRAMA CONSERVACIÓN EX-SITU/ANTONIO RIVAS (*cría de lince*)

EL CENTRO DE CRÍA de Zarza de Granadilla, en Cáceres, se inauguró en marzo de 2011 (*arriba*). Este y otros centros están aportando los linces que se reintroducirán en zonas de donde la especie había desaparecido. Hispano fue un ejemplar liberado en 2012 en la zona de Gaudalmellato (Córdoba) en el marco del proyecto Life Iberlince (*derecha*).

la vulnerabilidad a cepas víricas poco agresivas (como la que causó el brote de leucemia felina que provocó una alta mortalidad en Doñana en 2006) y la elevada tasa de mortalidad no traumática parecen indicar que los linces presentan una baja competencia inmunitaria. También muestran una incidencia especialmente alta de otros trastornos que podrían tener una base genética aún sin identificar, como el criptorquidismo (cuando uno o ambos testículos no descienden), la depleción linfoide (déficit de linfocitos en la sangre) y la epilepsia juvenil.

Asimismo, en la población más endogámica (Doñana) observamos una tendencia hacia una disminución del tamaño de la camada (el número de crías que nacen en un parto) y un incremento de la mortalidad no traumática (por enfermedades de distinto tipo) en el período 2002-2008 respecto a los del período 1983-1998, coincidiendo con una pérdida de diversidad genética y pese a la aplicación de medidas de conservación intensivas, incluida la alimentación suplementaria. La concurrencia del deterioro genético, por un lado, y de la reducción de la capacidad reproductora y la supervivencia, por otro, hace pensar que la especie, y en particular la población de Doñana, se hallaba a inicios de los 2000 en una dinámica que la estaba abocando a la extinción.

Las repercusiones de la erosión genética analizadas hasta aquí hacen referencia a la pérdida de eficacia biológica (capacidad reproductora y supervivencia) de los individuos y, por tanto, tienen lugar a corto o medio plazo. Por el contrario, el efecto de la pérdida de adaptabilidad de la especie al entorno como consecuencia de su escasa diversidad genética es mucho más difícil de predecir, puesto que este tipo de procesos se dan a largo plazo. No obstante, cabe pensar que la diversidad genética tendrá un papel crucial en la supervivencia de la especie en los próximos siglos, en los que se verá si tiene o no la capacidad de adaptarse a los cambios ambientales que se hallan hoy en marcha.

¿QUÉ PODEMOS HACER?

Al conocer que el lince sufría una elevada erosión genética y que esta podía estar perjudicando la eficacia biológica de los individuos, nos planteamos si existirían medidas que pudieran aliviar la situación. El comienzo de las actuaciones de gestión a principios de este siglo, incluidos los programas de cría en cautividad y de conservación de las poblaciones silvestres, nos ofrecieron una oportunidad única de aplicar estrategias de gestión genética para tratar de mejorar las oportunidades de persistencia del lince ibérico a medio y largo plazo.

Las dos poblaciones que aún sobrevivían en el año 2000 (Doñana y Sierra Morena) presentaban no solo un patrón genético muy acusado de baja diversidad y alta consanguinidad, como ya hemos comentado, sino también una elevada diferenciación genética entre una y otra. La primera decisión importante que había que tomar era si las dos poblaciones debían gestionarse de manera independiente o debían mezclarse y tratarse como una única unidad. Básicamente, ello significaba estimar si los riesgos de depresión por exogamia (pérdida de eficacia por mezcla de poblaciones genéticamente diferenciadas) en la gestión conjunta resultarían superiores a los de depresión por endogamia derivadas de la gestión independiente. Dicho de otro modo, había que sopesar si era más arriesgado cruzar que no cruzar individuos de las distintas poblaciones. Ambas son opciones legítimas y tienen sus pros y contras, pero en la decisión debimos tener en cuenta la dinámica evolutiva que había dado lugar a la diferenciación genética actual. Si esta se había producido como resultado de la acumulación de diferencias a lo largo de un período prolongado de aislamiento, quizás las poblaciones habían desarrollado sendas adaptaciones al entorno local. De ser así, la hibridación podría eliminar tales adaptaciones y disminuir la viabilidad de ambas poblaciones. Por el contrario, podía suceder que las diferencias se debieran a la contracción demográfica y a la fluctuación azarosa de las frecuencias alélicas que esta conlleva, con una eventual pérdida aleatoria de variantes y una acumulación de variantes deletéreas distintas en cada población generadas por la deriva genética. En este otro caso, cruzar las dos poblaciones ayudaría a reducir la consanguinidad y a aumentar la diversidad, puesto que los individuos híbridos serían muy poco homocigotos (consanguíneos) y la población mezcla reuniría las variantes potencialmente adaptativas que persistieron en cada una de ellas.

En el caso del lince ibérico, los análisis de la variación histórica demostraban que la pérdida de diversidad y la diferenciación

GRUPO TRAGSA (*centro de cría*); FRANCISCO VILLAESPESA, PROGRAMA DE CONSERVACIÓN EX-SITU DEL LINCE IBÉRICO (*liberación de Hispano*)

entre las dos poblaciones era reciente y se debían predominantemente a la acción de la deriva genética, lo que respaldaba la gestión conjunta de ambas. Ello se concretó, por un lado, con la cría en cautividad de una población mixta generada a partir de las dos poblaciones y, por otro, con el traslado de individuos de una población a la otra. Esta última medida se inició en 2007, con la liberación en Doñana de un macho nacido en Sierra Morena. A partir de 2009 se realizaron reintroducciones de individuos criados en cautividad en cuatro zonas de donde el lince había desaparecido: inicialmente, en Guadalmellato y Guarrizas (Andalucía), en el marco del proyecto Life 2006-2011; y después, en el valle del Matachel (Extremadura), Montes de Toledo y Sierra Morena oriental (Castilla-La Mancha) y el Parque Natural del Valle del Guadiana (Portugal), como parte del proyecto Life 2011-2016.

La gestión implementada hasta la fecha ha conseguido aumentar la diversidad genética de las poblaciones cautivas, reintroducidas y remanentes, lo que parece estar teniendo consecuencias positivas sobre la reproducción y la supervivencia de los individuos. Así lo sugiere el aparente éxito de los individuos mixtos, tanto en Doñana, actualmente compuesta en una alta proporción por los descendientes del primer macho trasladado desde Sierra Morena, como en cautividad, donde los individuos mixtos parecen tener una reproducción y supervivencia algo mayor, y en ningún caso menor, que los puros.

Una vez aplicadas estas medidas, debería evitarse que se produjeran nuevas pérdidas de diversidad y la acumulación de consanguinidad. La mejor forma de prevenirlo consistiría en dejar que las poblaciones crecieran rápidamente hasta que alcanzaran un tamaño «seguro», en el que la deriva genética ya no representara un peligro ni a corto ni a largo plazo. Estos tamaños se han estimado en unos 50 y 500 individuos, respectivamente, para el corto y el largo plazo.

Esas cifras resultarían válidas para una situación ideal en la que existiera un número igual de hembras y machos y todos los individuos se reprodujeran por igual, lo que los genetistas denominamos «tamaño poblacional efectivo». Pero en el mundo real no sucede así, por lo que se necesitan más individuos para alcanzar un tamaño de población seguro. El tamaño total (censal) de una población en un momento dado suele corresponder a un valor entre 5 y 10 veces superior al que es su tamaño efectivo (aunque esta proporción varía mucho entre especies). De este modo, cuando, a inicios de los años 2000, los tamaños censales de las poblaciones remanentes de Doñana y Sierra Morena eran de unos 50 y 100 individuos, respectivamente, los tamaños efectivos estimados correspondían a 10 y 20 individuos, unas cifras muy bajas e insuficientes para garantizar la viabilidad de las poblaciones. Según esto, el tamaño censal para el lince ibérico debería ser entre 250 y 500 individuos para evitar problemas a corto plazo, y entre 2500 y 5000 para mantener potencial adaptativo a largo plazo.

Sin embargo, en teoría podríamos obtener un mayor tamaño efectivo para un tamaño censal dado si gestionáramos la reproducción de manera que todos los individuos se reprodujeran por igual y, aún mejor, lo hicieran en función de lo singular que fuera su composición genética. Con esta estrategia se priorizaría la reproducción de los individuos menos emparentados genéticamente con el resto de la población, lo que llevaría a minimizar la consanguinidad promedio y a maximizar la diversidad genética. Por supuesto, la selección de emparejamientos solo es posible en cautividad, pero el mismo principio puede aplicarse a la gestión de los traslados de individuos de una población a otra y las reintroducciones (en zonas donde la especie había desaparecido) a partir de individuos en cautividad. Ello puede conseguirse si los animales que se van a trasladar o liberar se seleccionan en función de su parentesco con los residentes. Estas medidas de gestión genética las venimos implementando sistemáticamente en la población cautiva y para dirigir las liberaciones realizadas hasta la fecha, sobre la base de la información aportada por los marcadores moleculares y por la genealogía cuando esta se conoce.

RETOS PENDIENTES

A pesar del éxito de las medidas llevadas a cabo, la aplicación de la genómica a la gestión podría ir un paso más allá si conseguimos identificar y controlar aquellas variantes genéticas deletéreas que más están limitando la reproducción y la supervivencia. La actual gestión, orientada a maximizar la diversidad genética, tiende a homogeneizar las frecuencias de todas las variantes existentes en la población. Tal estrategia favorece también a las variantes deletéreas, puesto que en las poblaciones pequeñas no han sido purgadas eficientemente por la selección natural. La disponibilidad de marcadores asociados a defectos genéticos, como la criptorquidia o la epilepsia juvenil, posibilitaría una gestión que conciliara el mantenimiento de la máxima diversidad con la disminución de la incidencia de estas enfermedades.

Los planes de conservación de especies amenazadas requieren un enfoque multidisciplinar que integre todos los aspectos que pueden limitar la viabilidad tanto a corto como a largo plazo. El lince ibérico es un caso paradigmático que ilustra no solo los problemas genéticos, sino también sus posibles soluciones. Las medidas aplicadas han aumentado la diversidad genética de las poblaciones en libertad, especialmente en la población de Doñana. Pero todavía debe hacerse frente a varios retos para garantizar la persistencia del lince. Desde el punto de vista genético, estos consisten en extender el seguimiento y la gestión genética a las nuevas poblaciones para llegar a una gestión integral de la especie, y dar el salto al uso de herramientas genómicas, las ya disponibles y las que puedan desarrollarse en el futuro. En este proceso, el lince ibérico servirá también como una piedra de toque para evaluar el impacto real de la genómica en la conservación de las especies.

PARA SABER MÁS

Genetics at the verge of extinction: insights from the Iberian lynx. M. Casas-Marce et al. en *Molecular Ecology*, vol. 22, n.º 22, págs. 5503-5515, 2013.

Extreme genomic erosion after recurrent demographic bottlenecks in the highly endangered Iberian lynx. F. Abascal et al. en *Genome Biology*, vol. 17, n.º 1, pág. 251, 2016.

Novel efficient genome-wide SNP panels for the conservation of the highly endangered Iberian lynx. D. Kleinman-Ruiz et al. en *BMC Genomics*, vol. 18, n.º 1, pág. 556, 2017.

Spatiotemporal dynamics of genetic variation in the Iberian lynx along its path to extinction reconstructed with ancient DNA. M. Casas-Marcé et al. en *Molecular Biology and Evolution*, vol. 34, n.º 11, págs. 2893-2907, 2017.

Proyecto Life+Iberlince. «Recuperación de la distribución histórica de lince ibérico (*Lynx pardinus*) en España y Portugal». **iberlince.eu/index.php/esp/**

Programa de conservación ex-situ del lince ibérico. **lynxexsitu.es/index.php**

EN NUESTRO ARCHIVO

El lince ibérico. Progenie molecular. Juan F. Beltrán, John E. Rice y Rodney L. Honeycutt en *IyC*, mayo de 1996.

Lince ibérico. Genética y distribución. Francisco Palomares en *IyC*, agosto de 2004.

El jaguar: genes y conservación. Manuel Ruiz García en *IyC*, junio de 2013.

Revista: Revolución Molecular en las Ciencias Naturales

Biología molecular en la conservación del lince ibérico

La conservación de la biodiversidad en un escenario de cambio global acelerado constituye uno de los grandes retos de la humanidad para las próximas décadas.

José A. Godoy

Elena Marmesat

María Lucena Pérez

Daniel Kleinman Ruiz

Departamento de Ecología Integrativa. Estación Biológica de Doñana, CSIC

La genética ha pasado a ser un elemento fundamental del enfoque multidisciplinar que esta tarea requiere, dando origen al área de conocimiento de la genética de la conservación.

El papel de la genética en conservación tiene una doble dimensión. Por un lado, las especies amenazadas sufren cambios en su composición genética y estos cambios pueden afectar a la supervivencia y a la reproducción retroalimentando una espiral de declive que podría conducir a la extinción, incluso si las causas originales del declive desaparecen. Lo

más preocupante a corto plazo son los efectos nocivos de la consanguinidad y, a más largo plazo, la pérdida de adaptabilidad asociada a una escasa diversidad genética. Conocer bien los procesos que llevan a este deterioro genético y aportar métodos eficaces para la evaluación y gestión de estos problemas son objetivos fundamentales de la genética de la conservación. Por otro lado, y en un aspecto mucho más pragmático, distintos tipos de marcadores moleculares han resultado muy eficaces para el seguimiento y la gestión de especies amenazadas.

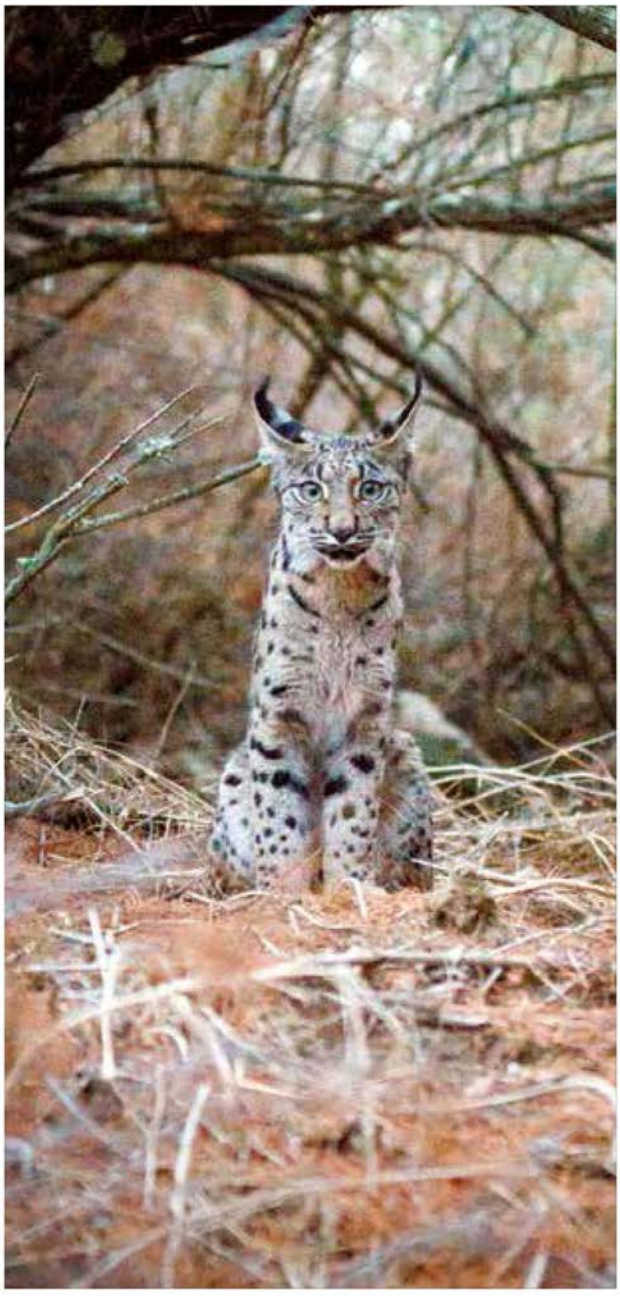


Figura 1
Cachorro de lince ibérico de la población de la cohorte de 2017 en Doñana. Foto de Iñigo Sánchez.

El lince ibérico (*Lynx pardinus*) es un excelente ejemplo de una especie que ha llegado al borde de la extinción; fatal desenlace que se ha evitado *in extremis* gracias a intensos esfuerzos de conservación. Tras el intenso declive sufrido durante el s. XX, en 2002 apenas quedaban unos 100 individuos repartidos en dos poblaciones desconectadas entre sí (Doñana y Andújar-Cardeña). Hoy día, se estiman más de 400 individuos de vida libre repartidos en las dos poblaciones remanentes y cinco sitios de reintroducción. El intenso declive y el seguimiento y la gestión activa a la que ha sido sometida la especie la han convertido en un excelente modelo de estudio para la genética de la conservación.

Aquí repasamos la contribución que ha realizado la genética molecular al conocimiento y la conservación de esta especie emblemática, desde los primeros enfoques basados en unos pocos marcadores moleculares neutrales hasta la reciente incorporación de la genómica, siguiendo la tendencia histórica recogida en el capítulo introductorio de este dossier.

MARCADORES TRADICIONALES: ADN MITOCONDRIAL Y MICROSATÉLITES

Relaciones filogenéticas

La secuenciación de fragmentos del ADN mitocondrial ha iluminado en primer lugar las relaciones del lince ibérico con otros linces y otros felinos. Distintos estudios han identificado alternativamente al lince boreal y al lince canadiense, al ibérico y el canadiense o al ibérico y boreal como especies hermanas. La incorporación de secuencias nucleares inclinó finalmente la balanza hacia el lince boreal como especie hermana del ibérico y permitió datar la divergencia entre ellos en unos 1.09 o 1.18 millones de años. Sin embargo, secuencias de genoma mitocondrial completo indican un antepasado común más reciente con lince canadiense, indicando un probable caso de clasificación incompleta de linajes para este genoma durante la divergencia de estas tres especies en un intervalo relativamente corto de tiempo (1.1-1.6 millones de años).

Genética poblacional: diversidad, consanguinidad y estructura

El análisis de 36 marcadores de tipo microsatélite a nivel poblacional permitió constatar un estado genético preocupante para la especie a principios del siglo XXI. Las dos únicas poblaciones remanentes estaban muy empobrecidas genéticamente y mostraban altos niveles de consanguinidad, en especial la población de Doñana, y contenían acervos genéticos altamente diferenciados. La sospecha inicial de que estos patrones eran consecuencia directa del declive sufrido durante el s. XX, y por tanto de la acción de la deriva genética, se vio en parte confirmada por los análisis de muestras históricas representadas en colecciones públicas y privadas. No obstante, estos análisis también revelaron que el deterioro y la estructuración genética estaban ya avanzados en las épocas cubiertas por estas muestras (1856–1990), apuntando a un declive temprano que diversos tipos de datos fechan en torno a hace tres o cuatro siglos. Al mismo tiempo, las secuencias de mitogenoma muestran una ausencia completa de diversidad a nivel de proteínas en el

presente. Por el contrario, las secuencias obtenidas de muestras arqueológicas y paleontológicas (43.500–2070 años de antigüedad) muestran una mayor pero ya limitada diversidad genética y sugieren una ausencia de diferenciación a escala peninsular en un pasado más remoto. El estado genético actual del lince ibérico parece pues ser la consecuencia de bajos tamaños poblacionales durante la mayor parte de la historia de la especie, declives sufridos hace varios siglos y un efecto dramático del declive del siglo XX.

Herramientas para el seguimiento y la gestión

Los marcadores tradicionales han aportado también herramientas eficaces para el seguimiento y la gestión de la especie, incluyendo la identificación de restos a nivel de especie e individuo.

Las secuencias de ADN mitocondrial obtenidas de restos biológicos muestreados en campo de manera no invasiva (excrementos, fundamentalmente) pueden identificar de manera no ambigua la especie de la que proceden y aportar así evidencias de su presencia en lugares concretos. Con el objetivo de aplicar este enfoque al análisis de la distribución global de la especie, pero eliminando los costes asociados a la secuenciación de ADN, desarrollamos un ensayo basado en una PCR diagnóstica de lince ibérico. La aplicación de este ensayo a un muestreo extensivo de las zonas de presencia histórica en Andalucía identificó a las poblaciones de Doñana y Andújar-Cardeña como las únicas poblaciones de lince ibérico remanentes en el año 2000. A pesar de informes de presencia de lince en otros lugares, estas poblaciones se han mantenido como las únicas poblaciones reproductoras confirmadas a nivel global hasta el inicio de las actuaciones de reintroducción.

El genotipado de marcadores de tipo microsatélite, junto con marcadores moleculares de sexo, permite el seguimiento de la especie a nivel individual por mé- todos no invasivos. Aplicados de manera sistemática en un área determinada pueden generar un censo poblacional, proporcionar estimas de tamaño poblacional, delimitar territorios individuales, identificar inmigrantes y generar una genealogía de la población que puede utilizarse para la gestión genética. La asignación molecular de parentales ha permitido reconstruir una genealogía bastante completa y de varias generaciones de la población de Doñana y constatar el éxito de los machos territoriales en asegurar la reproducción con hembras con territorio solapante.

Por último, los marcadores microsatélites han resultado también críticos para la gestión genética de la población cautiva. Esta gestión está dirigida a preservar el máximo de la diversidad representada inicialmente en los fundadores y evitar la acumulación de consanguinidad, y se basa en minimizar el parentesco promedio de la población. Los

genotipos moleculares han aportado estimas de parentesco entre los fundadores que, de otra forma, habrían sido ignoradas durante la gestión y habrían conducido a resultados subóptimos.

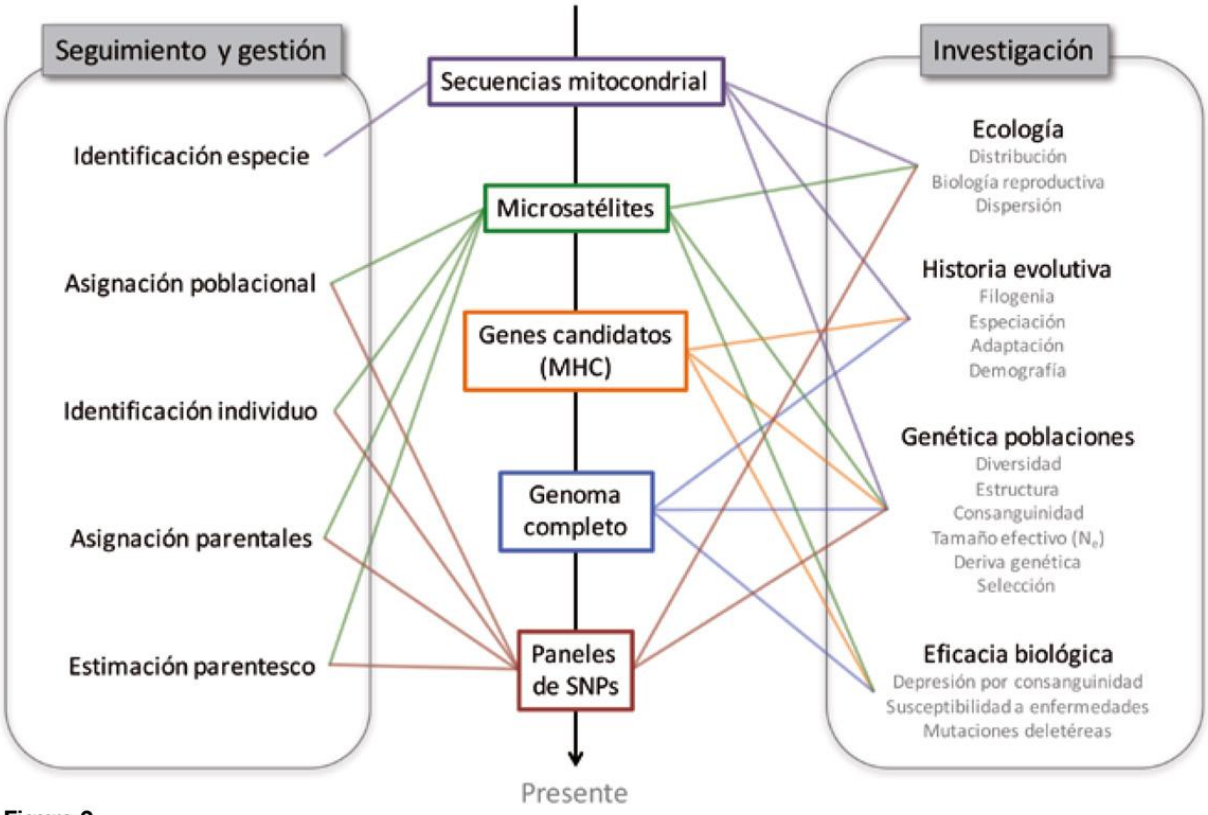


Figura 2
Usos de la información genética en la investigación y la conservación del lince ibérico. En el centro los distintos tipos de datos utilizados, en un orden cronológico aproximado de arriba (más antiguo) abajo (más reciente). A la izquierda se listan las aplicaciones útiles para el seguimiento y la gestión y a la derecha las áreas en las que esta información ha aportado nuevos conocimientos. Las aplicaciones en seguimiento y gestión se basan frecuentemente en el uso de muestras obtenidas de manera no invasiva como excrementos, mientras que las muestras de museo, arqueológicas y paleontológicas han contribuido al conocimiento de la historia demográfica y a evaluar el impacto de ésta sobre la genética actual.

GENES MHC: DIVERSIDAD FUNCIONAL Y RESPUESTA A ENFERMEDADES

Para trascender de la información demográfica que pueden darnos unos cuantos marcadores neutrales e investigar patrones adaptativos, las miradas se han vuelto hacia genes funcionales cuya diversidad es relevante para la supervivencia de los individuos. Estos genes están recibiendo atención también desde la genética de la conservación porque si pierden diversidad podría aumentar la susceptibilidad a enfermedades de las especies amenazadas y contribuir a su extinción. Pero también porque son claros candidatos a ser blanco de la selección balanceadora, fuerza evolutiva que podría contrarrestar la acción predominante de la deriva genética en especies en declive. Su análisis ha estado tradicionalmente limitado por las dificultades técnicas asociadas a la caracterización de una familia multigénica con un número variable de loci y múltiples

alelos por locus. Las nuevas tecnologías de secuenciación han aliviado algunas de estas dificultades, aunque persisten algunos problemas relacionados con sesgos de amplificación entre alelos.

En un primer análisis de la variación en genes MHC por secuenciación de productos de PCR clonados y aná- lisis de conformación SSCP, identificamos una casi total ausencia de diversidad genética en Doñana para genes MHC de clase II (dos únicos alelos correspondientes a dos genes, ambos fijados). Recientemente, aplicando una estrategia de amplificación mejorada y secuenciación masiva pudimos detectar complementos alélicos a nivel de especie para genes clase II y clase I de tamaño parecido y muy solapantes a los observados en el lince boreal, además de otros ejemplos de compartición de alelos entre las cuatro especies del género y con otros felinos. Encontramos que el declive reciente ha supuesto una pérdida de alelos a nivel de especie en estos genes clave, pero se ha mantenido la diversidad funcional representada en los dominios de unión a antígeno. Esto puede ser el resultado de la selección balanceadora manteniendo linajes de alelos divergentes, pese a la acción de la deriva. En familias multigénicas como estas, esto puede ocurrir tanto porque se favorezcan los heterocigotos en cada locus, como porque se mantengan linajes divergentes en distintos genes, aunque cada uno de estos sea completamente monomórfico. Los estudios en curso pretenden caracterizar la diversidad MHC a nivel individual y poblacional para testar estas hipótesis y calibrar la posible contribución de estos genes a la aparentemente alta susceptibilidad a enfermedades observada en esta especie.

GENOMAS COMPLETOS

Historia evolutiva y demográfica, adaptación, y genómica poblacional

La genética de la conservación no ha quedado al margen de la reciente expansión de la genómica en múltiples áreas de conocimiento. Los enfoques genómicos en conservación están comenzando a revelar nuevos aspectos acerca de cómo los genomas son transformados por efecto de la deriva genética, y a sopesar la posible interacción de esta con otros procesos evolutivos como la mutación, la selección natural y la recombinación. Al mismo tiempo, prometen aportar herramientas aún más eficaces para el seguimiento y la gestión.

El lince ibérico ha sido una de las primeras especies amenazadas en incorporarse al selecto club de especies con genomas completos secuenciados. La secuenciación, ensamblaje y anotación de un primer borrador del genoma ha puesto el poder de los nuevos enfoques genómicos al alcance de la especie. Los primeros análisis genómicos han aportado conocimientos antes inaccesibles sobre el origen y la historia de la especie y han ayudado a caracterizar con un nivel de detalle antes impensable su estado genético.

La comparación del genoma del lince ibérico con el del boreal ha acercado la fecha de divergencia entre las dos especies a los 300.000 años, y ha revelado una larga historia de intercambio genético entre ambas, un proceso de hibridación e introgresión que parece ser común entre felinos. Además, ha revelado historias demográficas relativamente paralelas de las dos especies desde su divergencia, con fluctuaciones a través de los ciclos glaciares pero con tamaños poblacionales menores en el lince ibérico. En esta especie evidenció también un repentino declive al 10% del tamaño anterior hace unos cuatro siglos, confirmando las inferencias obtenidas de microsatélites y ADN mitocondrial. En lince ibérico encontramos señales de selección positiva en genes relacionados con la audición y la visión, entre otros, que podrían indicar adaptaciones propias de esta especie.

Los datos de re-secuenciación de genomas completos de 10 linces adicionales revelaron al lince ibérico como la especie con menor diversidad genética entre las secuenciadas hasta la fecha, por debajo de otras especies también amenazadas como el guepardo, el leopardo de las nieves, el gorila de montaña, por ejemplo. Un aspecto especialmente relevante es que la escasa diversidad existente en zonas codificantes es en muy alta proporción debida a variantes no sinónimas, algunas de las cuales han llegado a fijarse. Estas variantes, a menudo deletéreas, se mantienen a raya en especies demográficamente abundantes por la acción de la selección purificadora, un proceso que en poblaciones pequeñas ve reducida su eficacia por el efecto predominante de la deriva genética. La expresión de estas variantes, en su mayoría recesivas, se vería incrementada por efecto de la consanguinidad, lo que podría estar generando depresión por consanguinidad, como la observada para calidad seminal. La ocurrencia de consanguinidad reciente se visualiza en forma de largos bloques de genomas en homocigosis en muchos de los linces secuenciados, estimándose niveles de consanguinidad individual promedios de 0.32 en Doñana y 0.16 en Andújar.

NUEVAS HERRAMIENTAS GENÓMICAS

La secuenciación del genoma del lince ibérico ha aportado importantes recursos para el seguimiento y la gestión. Quizás el de aplicación más directa ha sido un catálogo de 1,6 millones de variantes genómicas de nucleótido sencillo (SNP) en la especie. De entre estos, seleccionamos alrededor de 1.500 distantes entre sí y con alta diversidad y los genotipamos en más de 300 muestras de linces contemporáneos. Los datos permitieron seleccionar un conjunto de *ca*. 300 marcadores máximamente informativos y mínimamente ligados y paneles reducidos de SNPs para las distintas aplicaciones en seguimiento y gestión. Estos paneles superan en poder, eficiencia y bajo coste a los 36 microsatélites utilizados hasta la fecha y su implementación en un seguimiento genético de poblaciones silvestres y reintroducidas puede ayudar a prevenir futuras pérdidas de diversidad y a disminuir los altos niveles de endogamia acumulados durante décadas, maximizando así las probabilidades de persistencia de la especie.

Un reto importante para el futuro será la identificación de variantes asociadas a caracteres fenotípicos importantes para la supervivencia y reproducción, incluidas aquellas asociadas a caracteres deletéreos y enfermedades genéticas que pueden estar lastrando la recuperación, como la epilepsia juvenil en el lince ibérico o la condrodistrofia en los cóndores de California. Con ello se haría posible la incorporación de una "selección purificadora asistida" a la gestión genética integral, al menos para unos pocos caracteres de gran efecto.

PERSPECTIVAS

La aplicación de la genética en la conservación de especies es ya una realidad en casos paradigmáticos como el del lince ibérico, aunque todavía tiene que ser incorporada a otros muchos organismos. Las tecnologías genómicas, en particular la secuenciación de genomas completos, siguen estando fuera del alcance de la mayoría de especies amenazadas, pero las nuevas técnicas de secuenciación de fracciones del genoma (e.g. RAD-Seq o *Genotyping-By-Sequencing*) están ya sustituyendo a los marcadores genéticos tradicionales.

Más allá de las variaciones en la secuencia de nucleótidos, las aproximaciones futuras podrían incorporar otros aspectos hasta hora poco explorados del genoma, como la variación en número de copias, la expresión génica y su regulación, y los cambios epigenéticos, todos ellos con muy probables implicaciones para la viabilidad de especies amenazadas.

Detener el ritmo creciente de pérdida de biodiversidad exige cambios drásticos en nuestra forma de relacionarnos con el planeta. Mientras esto ocurre, los esfuerzos por evitar la extinción de especies como el lince ibérico y por asegurar su viabilidad a largo plazo van a beneficiarse de nuevos enfoques, en los que la biología molecular jugará sin duda un papel importante.

PARA LEER MÁS

Abascal F, Corvelo A, Cruz F, et al. (2016) Extreme genomic erosion after recurrent demographic bottlenecks in the highly endangered Iberian lynx. Genome Biology 17, 251.

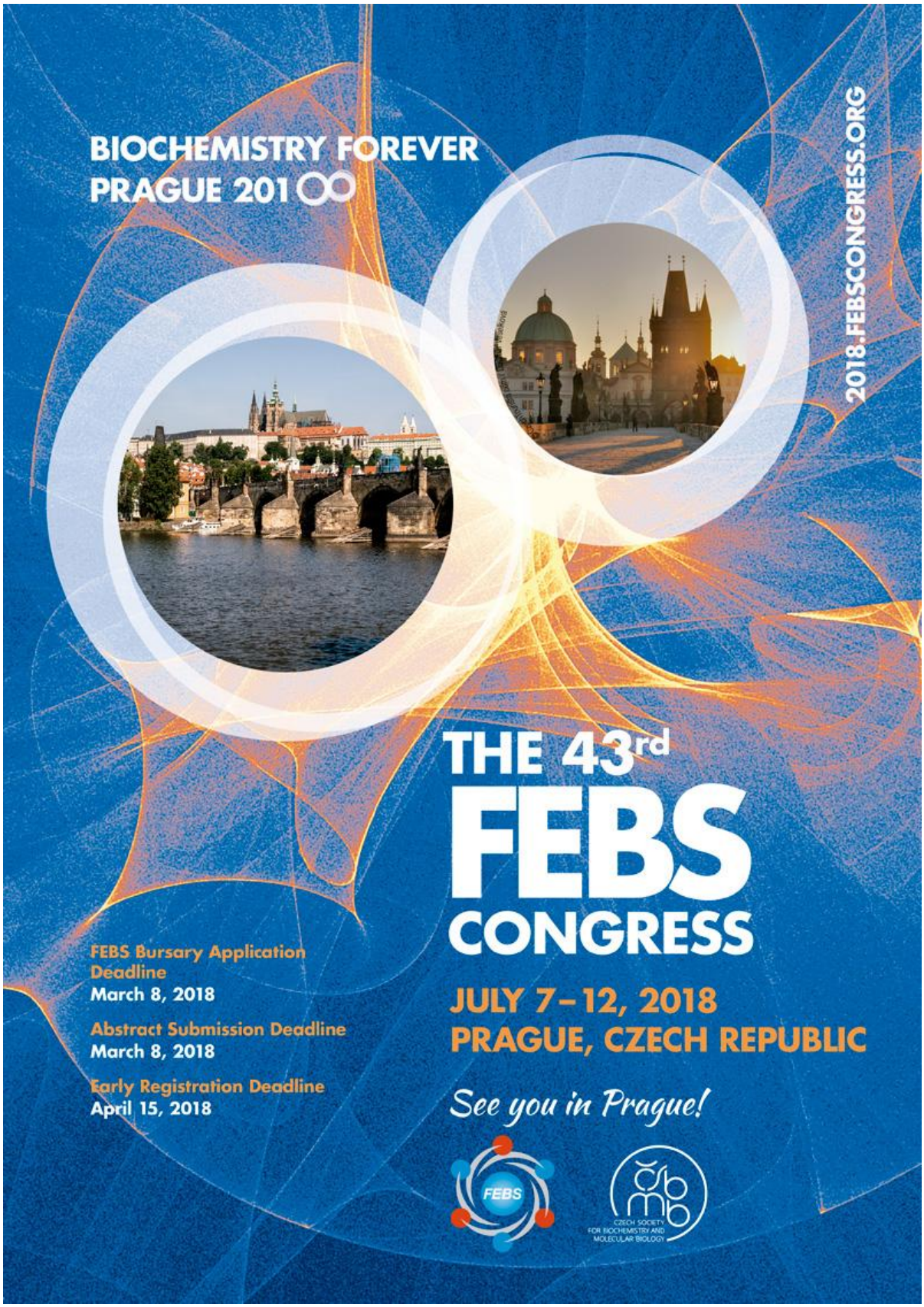
Casas-Marcé M, Marmesat E, Soriano L, et al. (2017) Spatiotemporal dynamics of genetic variation in the Iberian lynx along its path to extinction reconstructed with ancient DNA. Molecular Biology and Evolution en prensa.

Casas-Marce M, Soriano L, López-Bao JV, Godoy JA (2013) Genetics at the verge of extinction: insights from the Iberian lynx. Molecular Ecology 22, 5503-5515.

Kleinman-Ruiz D, Martínez-Cruz B, Soriano L, et al. (2017) Novel efficient genome-wide SNP panels for the conservation of the highly endangered Iberian lynx. BMC Genomics 18, 556.

Marmesat E, Schmidt K, Saveljev AP, Seryodkin IV, Godoy JA (2017) Retention of functional variation despite extreme genomic erosion: MHC allelic repertoires in the Lynx genus. BMC Evolutionary Biology 17, 158.

Palomares F, Lucena-Pérez M, López-Bao JV, Godoy JA (2017) Territoriality ensures paternity in a solitary carnivore mammal. Scientific Reports 7, 4494.



(https://2018.febscongress.org/)

(revista.php)

NEWSLETTER

¿Quieres recibir la revista en tu email?

Tu email aquí

No soy un robot

reCAPTCHA
Privacidad - Condiciones

ENVIAR

Spatiotemporal Dynamics of Genetic Variation in the Iberian Lynx along Its Path to Extinction Reconstructed with Ancient DNA

Mireia Casas-Marce,[†,1] Elena Marmesat,[†,1] Laura Soriano,[1] Begoña Martínez-Cruz,[1] Maria Lucena-Perez,[1] Francisco Nocete,[2] Antonio Rodríguez-Hidalgo,[3,4,5] Antoni Canals,[5,6,7] Jordi Nadal,[8] Cleia Detry,[9] Eloísa Bernáldez-Sánchez,[10] Carlos Fernández-Rodríguez,[11] Manuel Pérez-Ripoll,[12] Mathias Stiller,[13] Michael Hofreiter,[14] Alejandro Rodríguez,[15] Eloy Revilla,[15] Miguel Delibes,[15] and José A. Godoy*[,1]

[1]Department of Integrative Ecology, Doñana Biological Station (EBD) Spanish National Research Council (CSIC), Sevilla, Spain

[2]Grupo de Investigación MIDAS, Departamento Historia I (Prehistoria), Universidad de Huelva, Huelva, Spain

[3]Department of Prehistory, Complutense University, Madrid, Spain

[4]IDEA (Instituto de Evolución en África), Madrid, Spain

[5]Equipo de Investigación Primeros Pobladores de Extremadura, Casa de la Cultura Rodríguez Moñino, Cáceres, Spain

[6]IPHES, Institut Català de Paleoecologia Humana i Evolució Social (Catalan Institute of Human Paleoecology and Social Evolution), Tarragona, Spain

[7]Area de Prehistoria (Department of Prehistory), Universitat Rovira i Virgili (URV), Tarragona, Spain

[8]SERP, Departament de Prehistòria, Història Antiga i Arqueologia, Universitat de Barcelona, Barcelona, Spain

[9]UNIARQ – Centro de Arqueologia da Faculdade de Letras da Universidade de Lisboa, Alameda da Universidade, Lisboa, Portugal

[10]Laboratorio de Paleobiología, Instituto Andaluz del Patrimonio Histórico, Consejería de Cultura de la Junta de Andalucía, Seville, Spain

[11]Departamento de Historia, Faculdade de Filosofía y Letras, Universidad de León, León, Spain

[12]Prehistory and Archaeology Department, University of Valencia, Valencia, Spain

[13]Department of Ecology and Evolutionary Biology, University of California Santa Cruz, Santa Cruz, CA

[14]Evolutionary Adaptive Genomics, Institute of Biochemistry and Biology, Faculty of Mathematics and Life Sciences, University of Potsdam, Potsdam, Germany

[15]Department of Conservation Biology, Doñana Biological Station (EBD), Spanish National Research Council (CSIC), Sevilla, Spain

[†]These authors contributed equally to this work.

***Corresponding author:** E-mail: godoy@ebd.csic.es.

Associate editor: Michael Rosenberg

Bam files generated in this study have been deposited in the European Nucleotide Archive (study accession no. PRJEB15462). Mitogenomic consensus sequences for each sample have been deposited in the GenBank (accession nos. KX911255–KX911412).

Article

Abstract

There is the tendency to assume that endangered species have been both genetically and demographically healthier in the past, so that any genetic erosion observed today was caused by their recent decline. The Iberian lynx (*Lynx pardinus*) suffered a dramatic and continuous decline during the 20th century, and now shows extremely low genome- and species-wide genetic diversity among other signs of genomic erosion. We analyze ancient ($N = 10$), historical ($N = 245$), and contemporary ($N = 172$) samples with microsatellite and mitogenome data to reconstruct the species' demography and investigate patterns of genetic variation across space and time. Iberian lynx populations transitioned from low but significantly higher genetic diversity than today and shallow geographical differentiation millennia ago, through a structured metapopulation with varying levels of diversity during the last centuries, to two extremely genetically depauperate and differentiated remnant populations by 2002. The historical subpopulations show varying extents of genetic drift in relation to their recent size and time in isolation, but these do not predict whether the populations persisted or went finally extinct. In conclusion, current genetic patterns were mainly shaped by genetic drift, supporting the current admixture of the two genetic pools and calling for a comprehensive genetic management of the ongoing conservation program. This study illustrates how a retrospective analysis of demographic and genetic patterns of endangered species can shed light onto their evolutionary history and this, in turn, can inform conservation actions.

Key words: **Iberian lynx, ancient DNA, paleogenetics, genetic erosion, endangered species.**

Introduction

Species are becoming extinct at rates unprecedented in recent history as a consequence of human activity (Pimm et al. 2014; Ceballos et al. 2015). Populations of vertebrate species have decreased in size by an average of 52% in the last 40 years (McLellan 2014), and 15–40% of living species may have gone extinct by 2050 (Thomas and Williamson 2012). Habitat modification, fragmentation and destruction, pollution, climate change, overexploitation, and the spread of invasive species have been identified as the main drivers of the current biodiversity crisis. However, a hidden secondary threat in the form of genetic erosion often builds up as populations become small and isolated. Population genetics theory predicts that small and isolated populations progressively lose genetic diversity and accumulate genetic load as a consequence of genetic drift, and this may be exposed through inbreeding, resulting in inbreeding depression (Hedrick 2001). This makes them less able to adapt to environmental change and diminish their reproduction and survival. At the same time, population declines are often accompanied by fragmentation resulting in progressively reduced gene flow. Then, the above-mentioned processes operate independently in the resulting patches leading to increased genetic differentiation, unrelated to local adaptation (Frankham et al. 2002; Allendorf and Luikart 2007).

These processes have typically been assessed by comparing current genetic patterns among populations or species with different recent demographic history. However, current patterns of low genetic diversity and high differentiation can be the consequence of an alternative demographic and evolutionary history which entails contrasting risks of extinction: an equilibrium scenario derived from a long history of small population and restricted gene flow. This latter scenario entails a lower risk of inbreeding depression, since selection may have purged most deleterious variation. Besides, a genetic differentiation due to a stable migration-genetic drift equilibrium could be associated to local adaptation. The assessment of the relative weight of these two evolutionary histories is therefore of critical importance to evaluate the relative risks (inbreeding vs. outbreeding depression) and benefits of alternative management strategies (i.e., to mix or not to mix).

Fortunately, recent developments in the field of ancient DNA have allowed the characterization of genetic variation in past periods spanning decades—from museum specimens—to millennia—from archaeological and palaeontological samples. Historical and ancient genetic data can provide inferences of demographic history from the deep past throughout recent declines, and inform species conservation by providing landmarks of genetic diversity and differentiation statistics (Leonard 2008; Hofman et al. 2015). Furthermore, given a proper sampling of time periods and genetic markers, diachronic genetic studies can illuminate the dynamics of these processes, their dependence on demographic and geographical factors and their contribution to extinction.

The Iberian lynx (*Lynx pardinus*), because of its currently low genetic diversity and high differentiation, coupled with a fairly well-documented recent history of demographic decline and fragmentation, makes an outstanding model to characterize the genetic processes acting upon a species on its way to extinction. The species suffered a dramatic and continuous decline during the second half of the 20th century (Rodríguez and Delibes 2002; Ferreras et al. 2010; Palomares et al. 2011). With only two extant isolated populations—one located in the Doñana coastal plains and the other in the Eastern Sierra Morena mountains, ca. 240 km apart—and scarcely 100 individuals by 2002, the species was classified as critically endangered by IUCN (2002, 2006 and 2008 red lists; Rodríguez and Calzada 2015), and is generally recognized as the most endangered felid in the world (Nowell et al. 1996). Conservation actions implemented since then, including habitat protection, prey management, captive breeding, translocations and reintroductions have increased lynx numbers to more than 400 by 2015 (http://www.iberlince.eu/images/docs/3_InformesLIFE/Informe_Censo_2015.pdf; last accessed August 25, 2017), leading to its reclassification as "Endangered" in 2015 (Rodríguez and Calzada 2015).

Current Iberian lynx genetic diversity is low, especially in Doñana, and the two populations are highly differentiated (Johnson et al. 2004; Casas-Marce et al. 2013). Moreover, the Iberian lynx features one of the lowest genome- and species-wide diversity reported to date, as well as other signatures of genomic erosion, including high rates of potentially deleterious segregating and fixed mutations (Abascal et al. 2016). These patterns have been interpreted as the direct consequence of the species' recent decline (Johnson et al. 2004; Casas-Marce et al. 2013; Abascal et al. 2016). However, the complete lack of diversity in a short region of the mitochondrial genome in both historical and ancient samples has suggested that low-genetic diversity is an intrinsic feature of the species (Rodríguez et al. 2011). It remains thus unclear how much of the genetic impoverishment currently observed has been caused by recent versus long-term demography.

Here, we use the Iberian lynx as a model to characterize the micro-evolutionary processes operating on a species on the route to extinction. We compare current, historical, and ancient patterns of diversity and differentiation using nuclear microsatellite markers and mitogenomic sequences. We reconstruct past demography, quantify the relative contribution of the recent decline to current genomic erosion, assess the degree of accumulated genetic erosion in different populations and evaluate its relationship to past demography and to the final persistence or extinction of populations. Finally, we discuss the implications for the ongoing management and conservation programs.

Results

We extracted 230 modern tissue samples, and 296 historical samples from museum and private collection specimens (Casas-Marce et al. 2012) (fig. 1). We obtained good quality genotypes at 20 microsatellite markers for all the fresh samples and for 155 out of the 296 historical samples (52.4%). Genotyping success of historical samples varied among types of tissue in the same fashion as previously reported (Casas-Marce et al. 2010). Average allelic dropout rates were 1.7%

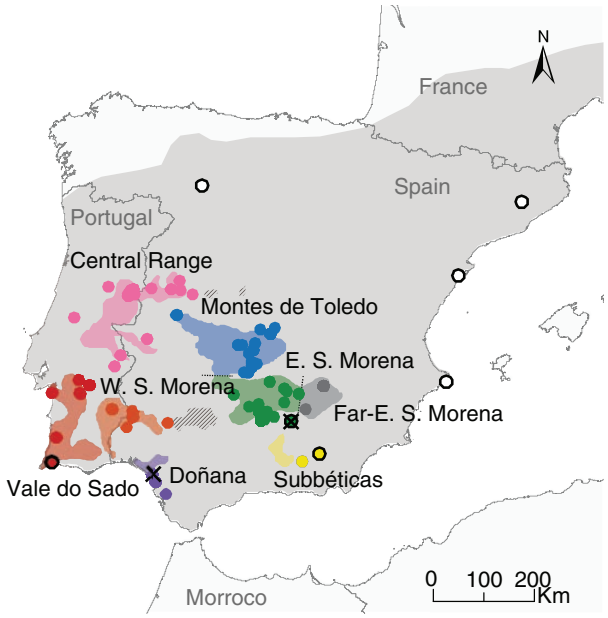


FIG. 1. Distribution of sampling across ancient and historical Iberian lynx ranges. Ancient range in light grey taken from Rodríguez and Delibes (2002). In Colour historical distribution according to country-wide surveys in the 1980s in Spain and 1989–1994 in Portugal, with populations delimited as in Rodríguez and Delibes (1992), except that we subdivided the largest Eastern Sierra Morena-Montes de Toledo population as suggested by genetic structure analyses. Points represent sampled localities, with outlined points corresponding to ancient samples and crosses respresenting contemporary samples; note that each point may represent several samples. Unsampled populations are shown in striped fill.

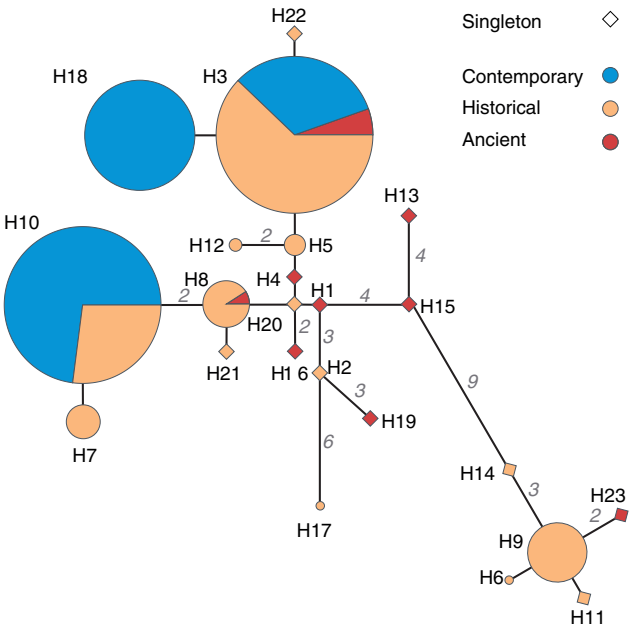


FIG. 2. Median-joining networks of mitogenomic haplotypes. Observed haplotypes are represented by circles whose sizes are proportional to the number of observations in each period. Haplotypes are connected by lines of length proportional to the number of mutations separating them (also indicated by small numbers). Haplotypes observed only once are depicted as diamonds to improve visibility.

(range 0–7.9%) and 2.0% (range 0–11.8%) across loci and samples, respectively, when calculated by comparing the replicates with the consensus genotypes, and 5% (range 0–25%) and 3.1% (range 0–30%) when comparing the 18 consensus genotypes obtained from historical samples with their respective genotypes from fresh samples. False alleles and Type 1 errors were 0.5% among loci and samples; other kinds of errors were not observed.

We also reconstructed whole mitochondrial genomes (16,449 bp) for a total of 158 Iberian lynxes: 10 ancient (dated 2.5–>43.5 ka), 83 historical (dating to 1700–1990), and 65 contemporary (1991–2010) samples, with each period containing samples distributed across the corresponding species range (supplementary tables S1–S4, Supplementary Material online). Overall, we observed 23 different haplotypes (fig. 2; supplementary fig. S9 and table S5, Supplementary Material online) defined by 40 variable sites, of which 39 are transitions and 1 is a transversion. Thirty-five variable sites occur in coding regions, of which eight are nonsynonymous variants (supplementary table S6, Supplementary Material online).

Demographic History

We used three different approaches to infer the demographic history of the species as a whole, and of each historical and contemporary population in different time periods. We first used a Bayesian Skyline Plot (BSP) (Drummond et al. 2012) based on whole mitochondrial genome data to explore long-term female effective population size. Then, an Approximate Bayesian computation (ABC) approach (Cornuet et al. 2014) was used with microsatellite data to gain insight into the more recent history spanning the last few centuries. Finally, we estimated census sizes from 1950 to 2015 and time of isolation for each historical population based on the available distribution and density data (Rodríguez and Delibes 2002).

BSP shows a nearly stable population of around 4,000 females throughout most of the covered lynx history, followed by a decline that started around 5,000 years ago and accelerated suddenly 400–450 years ago, when the population dramatically dropped from 2,000 to 20 females (supplementary fig. S1, Supplementary Material online). Next, we used ABC to adjust a model in which remnant populations diverged from an ancestral panmictic population formed by all other historical samples (supplementary fig. S2, Supplementary Material online). The divergence between Doñana and the rest of the historical population was estimated at the beginning of the 19th century (39.3 [31.3, 60.3] generations ago, which corresponds to an estimated decade of 1800s [1700s, 1840s]; mode [95% CI]), and its effective population size as 20 [14, 39] individuals (mode [95% CI]; considering a constant size over time; supplementary fig. S3*B*, Supplementary Material online). In the case of extant Eastern Sierra Morena, the estimated date is around 1950 [1880s, 1950s]) (11.1 [9.31, 23.3] generations ago), with a size of 29 [19, 56] individuals (supplementary fig. S3*B*, Supplementary Material online). These estimated isolation

dates, especially for Doñana population, are much older than what had been inferred by range reconstructions. Note that if we had used a model where divergence of populations was accompanied by gene flow or had allowed for larger population size in the past, the estimated divergence times would be even older, although the known demography and the results from other analyses (e.g., genetic structure) indicate that the current estimates are reasonable. On the other hand, the scarcity of samples from some areas and periods of historical populations impeded the consideration of further substructure for the ABC analysis. Although this could have led to an overestimation of the divergence times, the estimated dates are not far from those when Doñana and contemporary Eastern Sierra Morena start forming clusters separated from the historical population in STRUCTURE analyses (around 1900 or earlier for Doñana; between 1970 and 1990 for Eastern Sierra Morena; see *Genetic structure* section). Because the genetic differentiation must have postdated isolation by a few generations, the ABC estimations of population divergence dates do not seem highly overestimated if at all.

Finally, census sizes estimated for the historical populations over time based on records from 1950 to 2015 revealed that populations varied widely in size and trend. Two populations (Montes de Toledo or Eastern Sierra Morena) were relatively large by 1950 (above 750 individuals) but declined quite abruptly, while others (Far-E. Sierra Morena and Doñana) remained rather small (around 100 individuals) during the entire period (supplementary fig. S4, Supplementary Material online). All populations except Eastern Sierra Morena and Doñana were extinct by 2000, although the lack of data for the period 1985–2000 impeded estimation of the date of extirpation for most populations. The remnant Eastern Sierra Morena and Doñana reached 60 and 42 individuals, respectively, in 2000. Both populations increased their census sizes later following the adoption of conservation measures. In addition, the two larger populations remained connected until the mid 1980s, whereas most of the others became isolated at least 40 years earlier (figs. 3 and 4; supplementary fig. S4, Supplementary Material online). These two larger and connected populations (Montes de Toledo and Eastern Sierra Morena) are thus predicted to be the ones least affected by recent genetic drift and the best representation of the genetic variation of the species previous to the 20th century decline. Therefore, they are hereon considered central to the rest of the metapopulation.

Genetic Structure

We first analyzed different temporal and geographical partitions of microsatellite data with Factorial Correspondence Analyses (FCA) and with clustering algorithms implemented in STRUCTURE (Pritchard et al. 2000; Falush et al. 2007). The 3-D FCA plot shows a central cloud formed by the oldest samples from which more modern samples of the populations of extant Doñana, Eastern Sierra Morena and extinct Central Range and Far-E. Sierra Morena progressively separate through independent routes (supplementary fig. S5, Supplementary Material online).

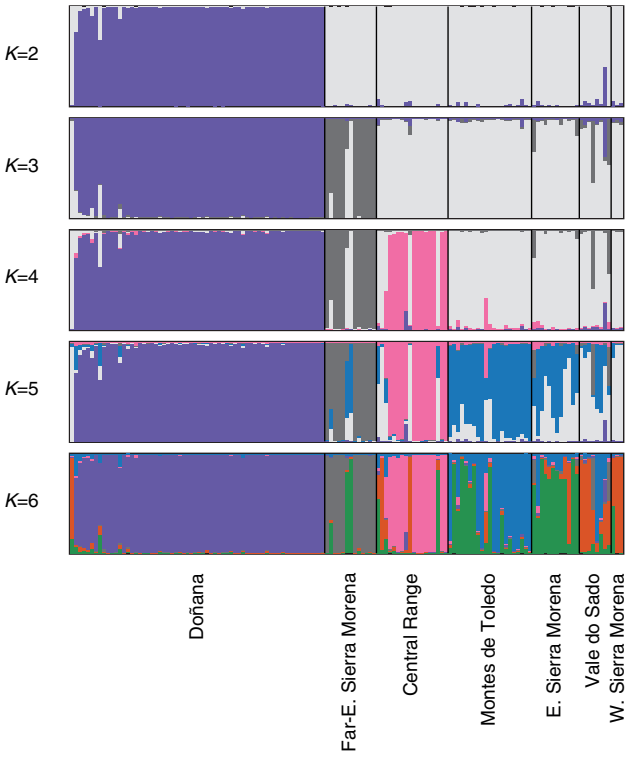


Fig. 3. Results of STRUCTURE analyses of historical microsatellite variation. Samples were first subdivided into two clusters, separating almost all Doñana samples from the rest ($K = 2$). A few older samples were partially assigned to the second historical cluster, with the oldest sample (dated in 1856) completely assigned to it. As K increases, other lynx populations are assigned to the new clusters, becoming differentiated from the rest: Eastern Sierra Morena ($K = 3$), Central Range ($K = 4$), Montes de Toledo-Eastern Sierra Morena ($K = 5$), and Western Sierra Morena-Vale do Sado ($K = 6$). Older samples from Eastern Sierra Morena and Montes de Toledo—two populations that have remained large and interconnected until the second half of the 20th century—are assigned to the same cluster (green), indicating that they were part of a single panmictic population that only recently became genetically differentiated. This genetic pool is probably the closest representation of the ancestral genetic variation of the species. See also supplementary figure S7, Supplementary Material online.

In the STRUCTURE analysis of all samples together (contemporary and historical), the modern samples of remnant populations are neatly separated from each other and from now extinct historical populations at $K = 3$, but older samples tend to cluster with the latter (supplementary fig. S6, Supplementary Material online). All individuals from Doñana are consistently grouped together in a separate Doñana cluster, but older samples are partially ($Q = 0.14$–0.27) and the oldest (1856) is completely assigned to the historical cluster ($Q > 0.95$). Similarly, most modern individuals (1970–2010) from Eastern Sierra Morena form a separate cluster, whereas older samples (1942–1973) show shared ancestry with the rest of historical samples from now extinct populations (supplementary fig. S6, Supplementary Material online). These results suggest that remnant populations became genetically differentiated from other historical populations between 1850 and 1900 in Doñana and between 1970

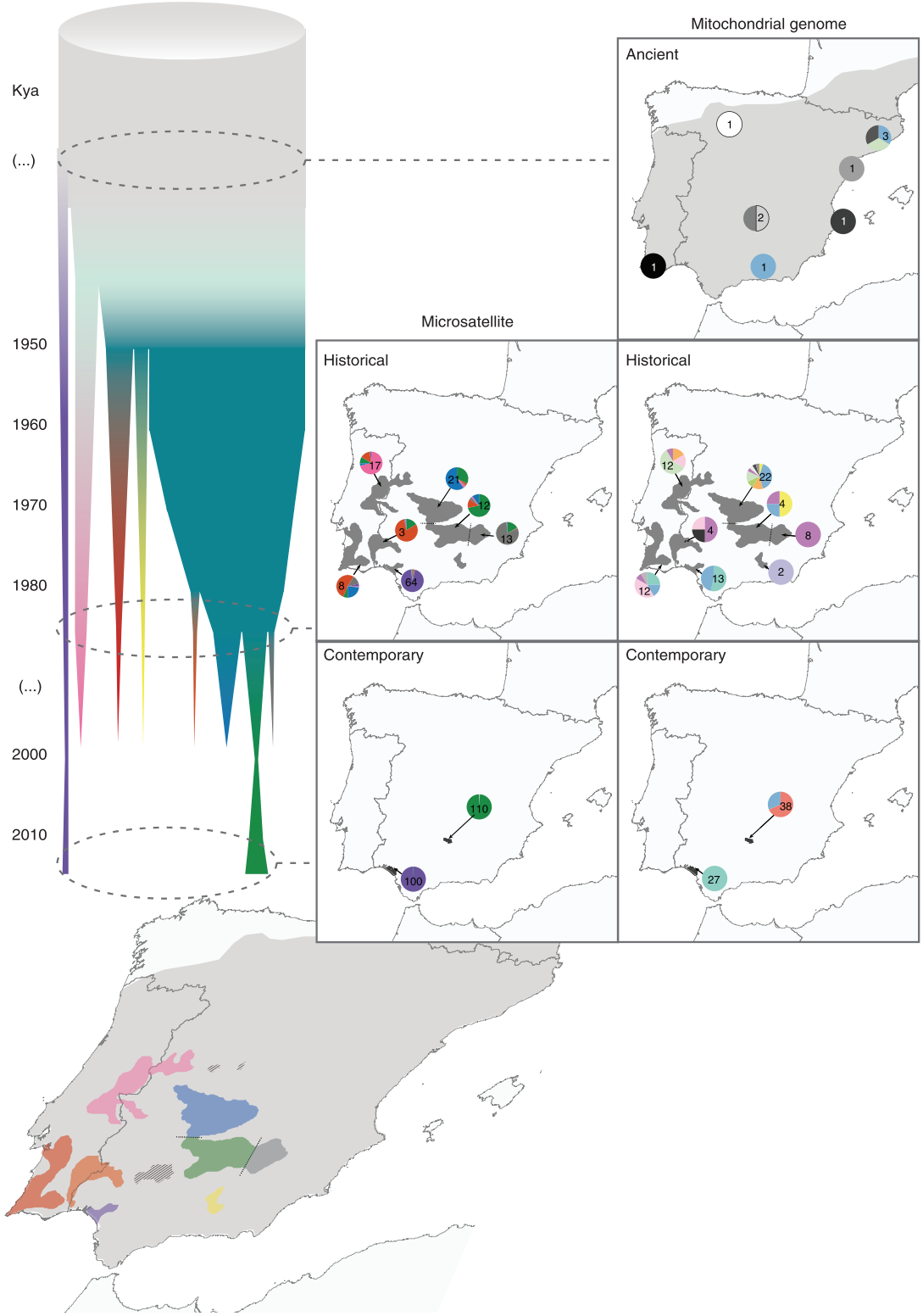


Fig. 4. Dynamics of population isolation and contraction, and genetic variation from ancient to contemporary times. The Iberian lynx population is represented by a cylinder projected on the distribution map, that becomes progressively fragmented into subpopulations which contract, become genetically differentiated and eventually go extinct. Maps represent the distribution of microsatellite (left) and mitogenomic variation (right) among ancient (top), historical (middle), and contemporary populations (bottom). Microsatellite pies represent the average coefficient of assignment to each of the clusters identified by STRUCTURE from historical ($K = 6$) and contemporary ($K = 2$) data sets (fig. 3 and supplementary fig. S7, Supplementary Material online). Mitogenome pies represent the distribution of mitochondrial genome haplotypes. Haplotypes observed only once are represented in shades of gray. Numbers within pies refer to sample size. See figure 2 and supplementary figure S9, Supplementary Material online, for networks depicting the relationship among haplotypes.

and 1990 in Eastern Sierra Morena, which is compatible with their respective dates of isolation estimated with ABC (supplementary figs. S2 and S3, Supplementary Material online).

The analyses of only historical samples (before 1990) of all populations confirmed the early and deep differentiation of Doñana ($K = 2$), and also revealed further population structure in the historical metapopulation (fig. 3; supplementary fig. S7*E–I*, Supplementary Material online). As K increases, new genetic clusters are identified that correspond to geographical populations defined a priori, in a sequence compatible with their estimated population size and dates of isolation (supplementary fig. S4, Supplementary Material online): Far-Eastern Sierra Morena ($K = 3$; unsuspected isolation), Central Range ($K = 4$; <1940), Montes de Toledo-Eastern Sierra Morena ($K = 5$; 1985), and Vale do Sado-Western Sierra Morena ($K = 6$; unknown date of isolation) (fig. 3). Genotypes tend thus to cluster by geographical location—even though some populations cover a wide temporal range (e.g., Doñana and Central Range; fig. 3; supplementary tables S1 and S2, Supplementary Material online). However, a temporal pattern is also evident in the data, as the oldest samples from each population tend to cluster together with other populations (fig. 3; supplementary fig. S7*E–I*, Supplementary Material online). For example, the oldest samples from Montes de Toledo and Eastern Sierra Morena are assigned to the same cluster. This indicates that the two adjacent populations were in fact one panmictic population that progressively became differentiated in recent times, in agreement with our estimates of times of isolation (1985; supplementary fig. S4, Supplementary Material online).

The genetic differentiation between pairs of populations is statistically significant and levels are in accordance with the expectations based on their geographical or temporal distance (supplementary fig. S8*A* and *B*, Supplementary Material online). The only nonsignificant pairwise comparison is Western Sierra Morena and the geographically contiguous Vale do Sado, represented by only three samples. The overall level of genetic differentiation is maximal between the two current populations ($F_{ST} = 0.419$) and lower among historical populations ($F_{ST} = 0.266$). Lowest differentiation is observed in the pairwise comparisons involving any of the two larger and more connected populations (Montes de Toledo or Eastern Sierra Morena), reaching a minimum for the comparison between them ($F_{ST} = 0.025$). As expected, comparisons among peripheral populations (Vale do Sado, Central Range and Far-Eastern Sierra Morena) show higher F_{ST} values, with the maximum recorded for Doñana versus Central Range ($F_{ST} = 0.405$). A similar pattern is captured with population-specific F values estimated by *2mod*, which may better reflect the differential action of genetic drift accumulated in each population. Temporal F_{ST} is relatively lower but also significant for comparisons of current versus historical samples of Doñana ($F_{ST} = 0.044$), and Eastern Sierra Morena ($F_{ST} = 0.071$), confirming the occurrence of changes in allele frequency through time in both localities (see also isolation-by-time analyses below).

Regarding mitogenomic variation, haplotype sharing is extensive in ancient samples, with the most extreme example being haplotype 3 occurring as far as in the northeast (Barcelona), southeast (Subbéticas), and southwest (Doñana) of the Iberian Peninsula (supplementary fig. S8*A* and *C*, Supplementary Material online). Insufficient sample sizes precluded the estimation of ancient mitogenomic F_{ST} values (fig. 4 and supplementary fig. S9, Supplementary Material online). Historical populations show moderate levels of geographical structuring with some haplotypes shared across populations ($F_{ST} = 0.39$; supplementary figs. S8 and S9, Supplementary Material online). Similar to microsatellite patterns, central populations are less differentiated than peripheral ones, including the peripheral Subbéticas population which was absent from the microsatellites data set (supplementary fig. S8*B* and *D*, Supplementary Material online). In contrast, current variation is totally structured in remnant Iberian lynx populations with nonoverlapping sets of mitogenomes, as previously observed based on short mitochondrial sequences (Casas-Marce et al. 2013) ($F_{ST} = 0.78$; fig. 4; supplementary fig. S9, Supplementary Material online).

Genetic Diversity

The comparison of current versus historical values of microsatellite and mitogenomic diversity allowed us to estimate the amount of diversity lost through time at the species level (fig. 5; tables 1 and 2). Overall, for microsatellites, expected heterozygosity (H_E) dropped moderately from 0.60 ± 0.13 in the historical period to 0.54 ± 0.13 at present (decrease to 90%; $S = 75.5$, $P = 0.003$), whereas allelic richness (A_R) dropped from 4.95 to 3.67 (74.1%; $S = 74.5$, $P < 0.001$) (fig. 5; table 1). A similar decrease was obtained for mitogenomic haplotype diversity ($Hd_{Historical} = 0.860 \pm 0.021 > Hd_{Contemporary} = 0.643 \pm 0.024$; 74.8%; $P < 10^{-11}$; fig. 5; table 2) and was most dramatic for nucleotide diversity ($\pi_{Historical} = 0.0005 \pm 0.00011 > \pi_{Contemporary} = 0.00018 \pm 0.00011$; 36%). A higher Hd was estimated for the ancient period ($Hd_{Ancient} = 0.978 \pm 0.054$), so that the comparison with current is significant ($P < 0.01$), but the comparison of historical is not, probably due to lack of power conferred by our relatively small ancient sample size ($N = 10$). Ancient nucleotide diversity is lower than that reported for the historical period ($\pi_{Ancient} = 0.00042 \pm 0.00025 < \pi_{Historical} = 0.00054 \pm 0.00028$); smaller sample sizes in the ancient and higher genetic structure in the historical data set may have contributed to this difference. The two extant Iberian lynx populations show extremely low current diversity, especially Doñana (fig. 5; tables 1 and 2; supplementary fig. S9, Supplementary Material online), for both microsatellites and mitogenomes. In particular, current overall mitogenomic diversity is the lowest reported for any mammal, with only three haplotypes defined by six positions, none of them translating to protein sequence variation (supplementary table S6, supplementary figs. S9 and 10, Supplementary Material online), but ancient and historical mitogenomic diversity, although higher, are still among the lowest for any mammal (supplementary figs. S9 and S10, Supplementary Material online).

At the population level, historical populations differ widely in their genetic diversity (tables 1 and 2). We found a general

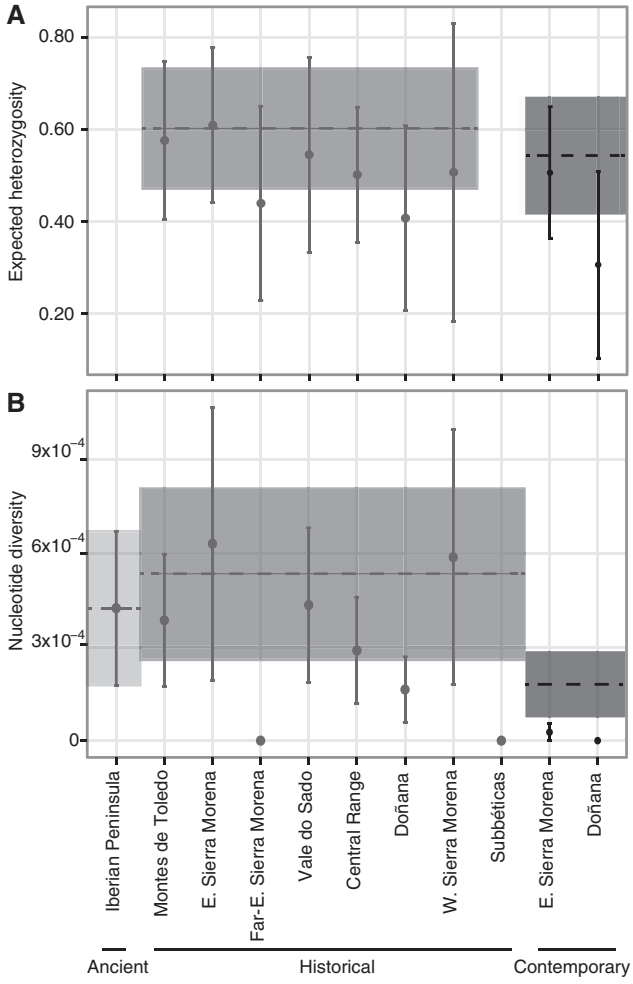


Fig. 5. Comparison of genetic diversity across periods. Microsatellite diversity was quantified as the unbiased expected heterozygosity (*A*), and whole mitochondrial diversity as nucleotide diversity (*B*). Points represent the average and standard error for each population, whereas the horizontal dashed lines and the corresponding shaded interval represent the same for the pooled samples of each period. Periods are color-coded in shades of grey. See tables 1 and 2 for other diversity and differentiation measures.

trend for higher diversity in central historical populations (e.g., Eastern Sierra Morena and Montes de Toledo) than in peripheral ones (Doñana, Far-E. Sierra Morena; fig. 5). For example, the peripheral Doñana population shows the lowest H_E and A_R while the central Eastern Sierra Morena and Montes de Toledo always rank as the top two. The microsatellite private allelic richness shows a different, but still parallel pattern. It is high in Montes de Toledo and historical Eastern Sierra Morena, but also in the otherwise derived Central Range, and is low in peripheral populations. This indicates that the peripheral harbour a subset of the diversity present in the central ones. Mitochondrial diversity indices vary similarly from complete uniformity in some of the peripheral populations (Far-Eastern Sierra Morena and Subéticas) to highest values in the central ones (table 2).

When we compare current versus historical genetic diversity within each of the two remnant populations, differences in diversity are also evident between periods. The contemporary population of Doñana harbours only 75% of the H_E ($S = 84.5$, $P = 0.0007$) and 83% of the A_R ($S = 103$, $P < 0.0001$) observed at earlier times in this same population. Similar patterns are observed for contemporary Eastern Sierra Morena with ratios of 89% for H_E ($S = 83.0$, $P = 0.001$) and 81.5% for A_R ($S = 82.0$, $P = 0.0012$) with respect to its historical variation. Therefore, both populations have progressively lost heterozygosity in the last decades, and they each represent a fraction of the overall historical diversity of the species (fig. 5; table 1).

We further explored the dynamics on genetic variation in each of the populations for which we had a wide temporal range. We plotted the standardized H_O of individuals through time and the relatedness of pairs of individuals relative to their temporal distance, the latter used as an indication of the intensity of allelic frequency fluctuations. The contrasting trends range from no changes in either diversity or allelic frequencies through time in Montes de Toledo, to the rather steep decrease and intense fluctuations in Doñana (fig. 6).

Discussion

Here, we reconstruct the genetic dynamics of the endangered Iberian lynx from ancient to contemporary times in order to shed light into the genetic processes that operated during the species decline toward its almost-extinction.

The Iberian lynx was once widespread in the Mediterranean biogeographic region of the Iberian Peninsula and may have reached Southern France and even Italy in the Pleistocene and the Holocene (Altuna 1972; Vigne and Pascal 2003; Yravedra 2005; Rodríguez-Varela et al. 2015). Taking into account this broad and continuous range, the most likely scenario for its ancestral genetic variation is a single genetically diverse and panmictic population. However, a limited pattern of isolation-by-distance could also have been possible depending on the extent of dispersal in the Iberian lynx. In fact, ancient mitogenomic variation was not highly structured, with haplotypes shared across the Iberian Peninsula, and overall mitogenomic diversity was probably higher than in historical times, but still low when compared with other mammals (fig. 4; supplementary figs. S9 and S10, Supplementary Material online).

Contrastingly, already by the early 20th century, both mitogenomic and nuclear markers revealed a structured metapopulation and locally low genetic diversity (figs. 3–5; supplementary fig. S9, Supplementary Material online). Unfortunately, we could not assess ancient microsatellite diversity, so we cannot be sure whether historical nuclear diversity was lower and structure higher than in more ancient times. A direct extrapolation of mitochondrial patterns is not warranted, because nuclear and mitogenomic patterns and dynamics might differ due to sex-biased dispersal and to their different sensitivity to bottlenecks (Fay and Wu 1999).

Historical genetic structure (fig. 3) could be either the consequence of an ongoing fragmentation and decline process that started earlier than previously thought, or the result of

Table 1. Microsatellite Diversity and Differentiation in Iberian Lynx Populations and Periods.

Epoch	Population	*N*	Dates Range	H_E	A_R ($n = 10$)	Private A_R	F_{IS}	*F* (2mod)	F_{st} Global
Current		210	1991–2010	0.54 (0.13)	3.67	0.05	0.27*		0.42
	E. Sierra Morena	102	1991–2010	0.51 (0.14)	2.65	0.02	0.01	0.61	
	Doñana	110	1991–2007	0.31 (0.20)	1.83	0.01	0.00	0.17	
Historical		143	1856–1990†	0.60 (0.13)	4.95	1.33	0.25*		0.27
	Montes de Toledo	22	1939–1977	0.58 (0.17)	3.02	0.11	0.04	0.07	
	E. Sierra Morena	10	1960–1990	0.61 (0.17)	3.25	0.13	0.14*	0.03	
	Far-E. Sierra Morena	13	1966–1989	0.44 (0.21)	2.42	0.08	0.11*	0.32	
	W. Sierra Morena	3	1970–1972	0.51 (0.32)			−0.02		
	Vale do Sado	8	1881–1956	0.55 (0.21)	2.80	0.03	0.22*	0.13	
	Central Range	18	1916–1993†	0.50 (0.15)	2.70	0.15	0.09*	0.19	
	Doñana	64	1856–1990	0.41 (0.20)	2.21	0.06	0.01	0.37	

NOTE.—H_E, expected heterozygosity; A_R, allelic richness; F_{IS}, population inbreeding coefficient; F_{2mod}, probability of identity by descent estimated by 2mod; F_{ST}, genetic differentiation coefficient.
*$p < 0.05$
†Includes a single sample with uncertain dating outside this range, around 1700.

Table 2. Mitogenomic Diversity and Differentiation in Iberian Lynx Populations and Periods.

Epoch	Population	*N*	Dates Range	Haplotypes	*S*	Hd (SD)	Pi (SD) (‰)	*K*	Tajima's *D*	Fu and Li's *F*	F_{ST} Global
Current		65	1991–2010	3	6	0.64 (0.02)	0.18 (0.11)	2.95	3.232***	2.15	0.78
	E. Sierra Morena	38	1991–2010	2	1	0.44 (0.06)	0.03 (0.03)	0.44	1.253	0.88	
	Doñana	27	1991–2007	1	0	0 (0)	0 (0)	0.00			
Historical		83	1881–1985	15	32	0.86 (0.02)	0.54 (0.28)	7.97	0.693	1.13	0.39
	Montes de Toledo	22	1939–1985	9	23	0.81 (0.07)	0.39 (0.21)	6.02	−0.183	0.69	
	E. Sierra Morena	4	1960–1972	3	19	0.83 (0.22)	0.63 (0.44)	9.83	−0.541	0.05	
	Far-E. S. Morena	8	1964–1971	1	0	0 (0)	0 (0)	0.00			
	W. Sierra Morena	4	1910–1972	3	17	0.83 (0.22)	0.59 (0.41)	9.00	−0.331	0.23	
	Vale do Sado	12	1881–1956	5	21	0.79 (0.09)	0.43 (0.25)	6.82	−0.099	1.13	
	Central Range	12	1881–1974	4	20	0.65 (0.13)	0.29 (0.17)	4.48	−1.459	−1.43	
	Doñana	13	1856–1982	2	5	0.54 (0.06)	0.16 (0.1)	2.69	2.392*	1.75	
	Subbéticas	2	1874–1880	1	0	0 (0)	0 (0)	0.00			
Ancient		10	43500–2070 ybp	9	24	0.98 (0.05)	0.42 (0.25)	6.64	−1.181	−1.12	

NOTE.—S, number of segregating sites; *Hd*, haplotype diversity; *Pi*, nucleotide diversity; *K*, average number of nucleotide differences; *D* and *F* are neutrality indices of Tajima (1989) and Fu and Li (1993), respectively.
*$p < 0.05$
*$p < 0.001$

the species' natural demographic dynamics. Given the evidence for predominance of recent genetic drift in local populations, we favour the view of an almost panmictic ancestral population, which started to contract and fragment before the sampled historical period. Indeed, mitogenomic data indicate the occurrence of a previous bottleneck around 400 years ago. Similarly dated bottlenecks were previously inferred for this species (Casas-Marce et al. 2013; Abascal et al. 2016), and other European carnivores (Breitenmoser 1998; Valdiosera et al. 2008; Schmidt et al. 2011), and are congruent with its contracted and patchy distribution in the period 1572–1897 inferred from historical records (Clavero and Delibes 2013). Although we can only speculate about the causes of this historical decline, they may well be related to an increase in anthropic pressures. The 16th century was a period of intense human population growth in Iberia and across Europe in general, and coincided with the extension of agriculture and the intensification of forest destruction, both of which may have initiated the decline of this Mediterranean shrubland specialist (Ellis et al. 2010). Both ancient and historical bottlenecks have impacted its genetic variation and are in large part responsible for the Iberian lynx being apparently the mammal species with the lowest genome- and species-wide diversity today (Abascal et al. 2016).

During the process of decline, the populations lost diversity and became genetically differentiated due to random fluctuations in allelic frequencies (fig. 6). Such scenario is supported by heterozygosity through time and isolation by time plots (fig. 6), as well as estimates of identity by descent (Column "F_{2mod}" in table 1). These results also reveal differences in the intensity of these processes across the Iberian lynx's range (fig. 6). The accumulated effects of genetic drift in each population are probably related to the dynamics of the fragmentation and decline process. Genetic drift started earlier and impacted to a greater extent the patches that became isolated sooner at the periphery of the species' range (fig. 4; Rodríguez and Delibes 2002). For instance, the Doñana population showed the strongest signal of genetic drift, consistent with its peripheral position and its long history of low effective size and genetic isolation, which we have estimated in around 20 individuals and lasting ca. 200 years (around 40 generations, at the very beginning of the 19th century;

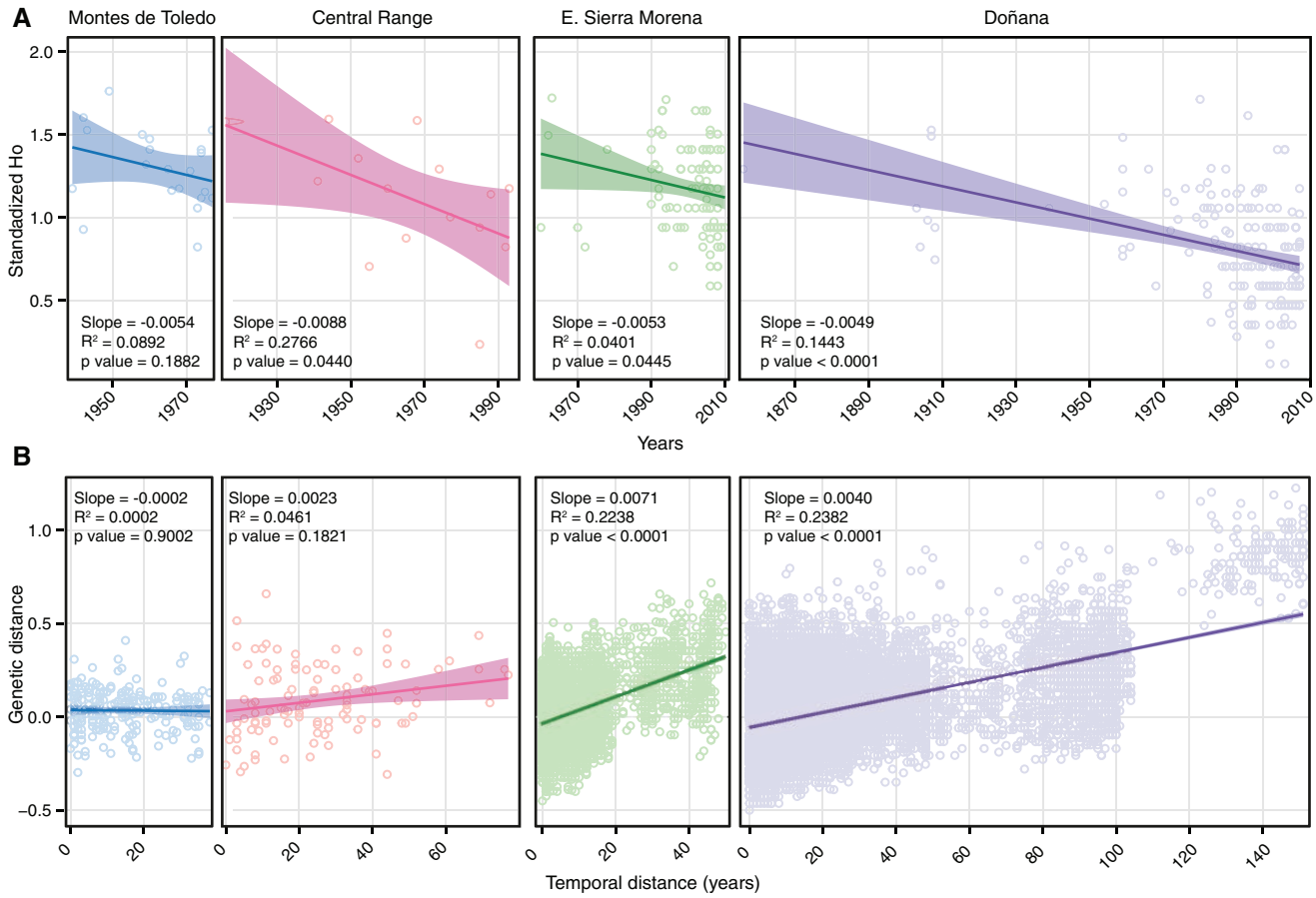


Fig. 6. Effects of genetic drift in Iberian lynx populations. The effect of drift is illustrated by the decrease in individual standardized observed heterozygosity (H_O) (*A*) and the increase of genetic similarity between pairs of individuals with time (*B*). The intensity of drift varies in the different populations in agreement with known demographic history, ranging from low in the large and connected Montes de Toledo population to high in the small and isolated Doñana population.

supplementary fig. S3*B*, Supplementary Material online). In contrast, we did not observe much genetic drift in Montes de Toledo, and only very recently in Eastern Sierra Morena. Both populations remained large and connected to each other, as well as to other more peripheral populations, until recently (fig. 4; supplementary fig. S3*B*, Supplementary Material online; Rodríguez and Delibes 2002). Thus, we hypothesize that they played a pivotal role in the metapopulation, acting as a reservoir of genetic diversity and as a source of gene flow to other populations. The major role of effective population size and gene flow in determining genetic patterns following decline and fragmentation is in line with expectations from population genetic theory and with empirical studies in other species (e.g., Méndez et al. 2014).

Despite extensive evidence showing how genetic erosion can negatively affect population viability (Frankham et al. 2010; Allendorf et al. 2013), the amount of drift accumulated in different Iberian lynx populations beared little regard to the final outcome of persistence or extinction (figs. 4 and 6). This study rather suggests that their final fate was generally determined by extrinsic factors. Possibly, the influence of genetics was overridden by increased extrinsic pressures in demographically and genetically healthy populations, and by the protection and active conservation granted to the most severely eroded population. Thus, while Central Range and Far-Eastern Sierra Morena populations did accumulate genetic erosion before their extinction by the late 20th century, the central population of Montes de Toledo remained largely unaffected by genetic drift until its final extirpation. Whereas in the former populations, genetic erosion might have contributed to population decline, this was unlikely to be the case for the latter. Montes de Toledo may have been abruptly pushed from a large and well-connected population to complete extirpation by deterministic factors acting in this area during most of the 20th century, namely landscape homogenization and lasting scarcity of prey (Rodríguez and Delibes 1990). On the other hand, we observe a striking persistence of the Doñana population despite a long history of small population size, isolation, and extreme genetic erosion, which according to recent studies may have affected reproduction and survival (Palomares et al. 2012). The persistence of Doñana could thus be attributed to the protection granted to the area by its designation as a royal hunting reserve during most of the 19th century and to the establishment of Doñana National Park in 1965.

While both remnant populations have progressively lost heterozygosity in the last decades (fig. 6), they have done so to very different extents. In contrast to the highly eroded

Doñana population, Eastern Sierra Morena represents the remnant of the large historical core population encompassing Montes de Toledo and Eastern Sierra Morena, and it is the current population genetically closest to the ancestral lynx population (fig. 4). Still, the contemporary diversities of both populations are lower than those reported for demographically healthy populations of its sister species, the Eurasian lynx (Casas-Marce et al. 2013; Ratkiewicz et al. 2014). However, the two contemporary populations conjunctly represent most of the former overall heterozygosity (90%), and also a moderate proportion (74.1%) of the allelic richness of the species in historical times, which were more in line with those observed in the Eurasian lynx (Ratkiewicz et al. 2014).

It must be noted that at least two sources of bias might have contributed to an underestimation of historical microsatellite diversity, and thus of the amount of diversity lost. First, allelic dropout in historical samples despite the use of replicates will cause heterozygotes to be called as homozygotes. This will reduce the observed population heterozygosity (H_O) more than H_E, and may thus contribute—together with spatial or temporal structure—to explaining part of the significant heterozygote deficit observed in most historical populations (table 1). Second, the microsatellites used were selected in previous studies (Johnson et al. 2004; Casas-Marce et al. 2013) to be highly variable in the current population, thereby introducing an ascertainment bias that would again lead to an underestimation of historical diversity and thus also of the amount of diversity recently lost.

In summary, our results show that the recent genetic erosion of Iberian lynx has been severe and has affected both microsatellite and mitogenomic diversity. Such erosion has three main components: (1) loss of diversity through time within populations, (2) an increasing differentiation between populations, and (3) extinction of genetically differentiated populations at the edges of the historical distribution.

Conservation Implications

Here, we demonstrate that the species' genetic makeup has been shaped by a long history of low population size and a sharp population decline around the 16th–17th century AD, well before its last well documented decline in the 20th century that left only two small remnant populations. However, we show that the extremely low genetic diversity of the two remnant populations (Casas-Marce et al. 2013; Abascal et al. 2016) is also the result of intense genetic drift and fragmentation occurring during the last decades. Due to the abrupt and recent nature of these changes, the risk of inbreeding depression in both remaining populations must be explicitly considered and addressed. Our results and conclusions are in contrast to those of a previous study reporting a long-term lack of mitochondrial diversity throughout the Iberian lynx's history based on short control region sequences, which led to the suggestion that the Iberian lynx's genetic diversity had always been low and was thus not a threat to its long-term viability (Rodríguez et al. 2011). Evidence for the occurrence of inbreeding depression in the species is accumulating in the form of heterozygosity-fitness correlations of sperm quality, reduced reproductive rates, increased nontraumatic mortality, and high rates of potentially genetic diseases (Peña et al. 2006; Jiménez et al. 2008; Palomares et al. 2012; Ruiz-López et al. 2012; Martínez et al. 2013). At the same time, the lower differentiation in historical and ancient times revealed from our data suggests that the currently observed genetic differentiation between the two remnant populations is mainly the result of genetic drift in recent times and cannot be attributed to independent adaptive evolution over long periods, indicating low risks of outbreeding depression (Frankham et al. 2011). Low-genetic diversity together with higher risks of inbreeding than outbreeding depression support the ongoing admixture of the two genetic pools both in captivity (Godoy et al. 2009) and in the wild through translocations (Simón et al. 2012). Although the positive contribution of admixture and other genetic management to Iberian lynx recovery has not yet been formally evaluated, the observation of an increased reproductive performance of admixed individuals suggests so (Godoy JA, unpublished data). Furthermore, genetic risks call for an intensive genetic monitoring and a sound comprehensive genetic management program for the species, which should also include the ongoing reintroductions. On the other hand, it is likely that the globally low genetic diversity will influence the chances of long-term persistence of the species, especially in a scenario of rapid global change (Fordham et al. 2013). Addressing the possibility of long-term reduced adaptive potential will have to carefully consider risks and benefits of less conventional measures to restore genetic diversity, like facilitated adaptation, genome edition, or assisted introgression (Thomas et al. 2013; Hamilton and Miller 2016).

Conclusions

Over time, species may have experienced complex demographic changes, including earlier anthropogenic impacts, which we cannot directly infer from contemporary genetic patterns and recorded demographic history. In the case of endangered species, disentangling the processes acting upon species throughout their history can help us understand how they became endangered and what we need to do to restore natural and healthy genetic patterns to guarantee their long-term survival. We show that the Iberian lynx has indeed lost a significant part of its already low genetic variation over time due to both recent and unsuspected older demographic declines, and that the contemporary pattern of high genetic differentiation between the remnant populations was caused by genetic drift during the last few centuries. Our retrospective look was only possible due to the availability of extensive ancient, historical, and contemporary samples (Casas-Marce et al. 2012), a careful selection of sampled tissues in historical specimens (Casas-Marce et al. 2010), recent technical advances in the field of ancient DNA (Hofreiter et al. 2015; Orlando et al. 2015), and the combination of nuclear markers and whole mitochondrial genomes.

To our knowledge, our analysis of the Iberian lynx is the most comprehensive retrospective analysis of genetic variation of an endangered species to date. It highlights how the evolutionary history of a species inferred solely from its current genomic makeup can be distorted, misinforming

management recommendations. Studying paradigmatic conservation cases like this one is also crucial for deepening our understanding of the demographic and genetic processes occurring during population declines, a much needed input for the effective restoration of endangered species and for the prospective assessment of species viability in a changing world.

Materials and Methods

Samples

We used a combination of modern, historical, and ancient samples (supplementary tables S1–S3, Supplementary Material online). Samples were grouped in a priori populations based on their origin or, when lacking, on results of clustering analyses. We also delimited periods based on recorded date of sampling, using the year 1990 as the limit separating contemporary and historical samples. This is approximately the time of the collapse of all but the two remnant Iberian lynx populations (Ferreras et al. 2010) and the time that roughly separates museum from fresh samples. We extracted DNA for a total of 230 fresh, 296 museum, and 58 archaeological samples.

Microsatellite Genotyping

Microsatellite markers were amplified in contemporary and historical samples. We used 20 microsatellites selected from 36 previously used for the analyses of modern Iberian lynx specimens (Casas-Marce et al. 2013) based on their good performance with degraded DNA. For historical samples we applied a preamplification multiplex approach (Piggott et al. 2004). We obtained good quality genotypes at 20 microsatellite markers for all the fresh samples and for 155 out of the 296 historical samples (52.4%).

Mitochondrial Genome Sequencing

For contemporary samples, we first used a long-range PCR approach to sequence the whole mitochondrial genome in 12 Iberian lynx (eight from Doñana and four from Eastern Sierra Morena). The amplified products were pooled equimolarly and individually tagged 454 sequencing libraries were prepared following the protocol described by Meyer et al. (2008a; 2008b). In addition, we used available whole-genome shotgun data for 31 different lynx (19 from Eastern Sierra Morena and 12 from Doñana), and shotgun reads from a separate capture-enrichment project targeting a subset of the nuclear genome in 34 individuals (13 from Eastern Sierra Morena and 21 from Doñana). We prepared a total of 111 historical and 20 ancient individually tagged libraries for Illumina or 454 sequencing following Meyer and Kircher (2010) or Maricic et al. (2010) (supplementary table S4, Supplementary Material online). *L. pardinus* biotinylated capture probes were prepared as described in Maricic et al. (2010) from the products of two overlapping long-rang PCR. Historical and ancient samples were enriched in target mitochondrial sequences by one or two consecutive rounds of hybridization–capture (supplementary table S4, Supplementary Material online). We mapped both merged and unmerged reads to the *L. pardinus* mitochondrial reference genome (Abascal et al. 2016) using BWA-mem (Li and Durbin 2009) with default parameters. Average coverage per sample ranged from 0.1× to 239.2× (supplementary table S4, Supplementary Material online). Only samples with total coverage over 5.2× were considered for further analyses. Before SNP calling we used *mapDamage* (Jonsson et al. 2013) on historical and ancient samples to track and quantify DNA damage patterns and rescale quality scores of likely damaged positions accordingly. Bam files generated in this study have been deposited in the European Nucleotide Archive under study accession number PRJEB15462.

Mitochondrial SNP Calling and Annotation

We simultaneously called SNPs on the pooled samples using *Freebayes* (Garrison and Marth 2012). We visually curated the 57 resulting variants by revising each position called as an SNP for all individuals to identify potential contaminations and artifacts. The number of reference positions not covered by any read was calculated for each individual using the command *genomecov* in *BEDTools* (Quinlan and Hall 2010). A consensus for each individual mitochondrial genome was constructed using the *FastaAlternateReferenceMaker* command in GATK (McKenna et al. 2010) (supplementary table S5, Supplementary Material online). Validated SNPs were annotated as genic/intergenic, transition/transversion, synonymous/nonsynonymous, and aminoacid change using Geneious v. 8 (http://www.geneious.com; last accessed August 25, 2017; Kearse et al. 2012) (supplementary table S6, Supplementary Material online).

Demographic Reconstruction Using Whole Mitochondrial Genomes

Past population dynamics was investigated with a BSP model using BEAST v1.8.1 (Drummond et al. 2012). We calibrated the tree using the divergence of the entire lynx lineage with a mean value of 2.52 My and a standard deviation of 0.4 My to estimate a substitution rate of 1.89×10^{-8} substitutions/site/year (1.16×10^{-8}–2.78×10^{-8} substitutions/site/year, 95% HPD). To construct the BSP, we assumed a HKY + G model of evolution and a strict molecular clock, and incorporated the information on the age of the sequences using the sampling date for contemporaneous samples, the date of death for the historical, and the radiocarbon estimated date for the ancient samples (supplementary tables S1–S3, Supplementary Material online).

Estimation of Divergence Parameters with ABC

We used an ABC approach based on coalescent simulations implemented in *DIYABC* version 2.1.0 (Cornuet et al. 2014) to estimate effective population sizes and divergence times between populations. We used a simple model that assumes constant population sizes and no gene flow between populations after divergence. We used a generation time of 5 years, grouped samples by decades and considered each decade to be separated from the next one by two generations (supplementary fig. S2, Supplementary Material online). Prior distributions for the effective population size of the historical population (N_1), contemporary Eastern Sierra Morena (N_2), and Doñana (N_3) were set based on previous estimates of

contemporary sizes (supplementary fig. S3*B*, Supplementary Material online; Casas-Marce et al. 2013). For the purpose of building the posterior distributions, we used as summary statistics the expected heterozygosity and the number of observed alleles for all samples in the analysis and pair-wise F_{ST} values between some of the samples. We ran 1,000,000 simulations and used the 1% closest to our data set to build the posterior distributions using a local linear regression technique (supplementary fig. S3*A*, Supplementary Material online; Beaumont et al. 2002).

Estimation of Historical Population Sizes and Dates of Isolation

During the period 1950–1990, Iberian lynx abundance was estimated from the textures of relative abundance published by Rodríguez and Delibes (2002) on a 10-km UTM grid. To estimate lynx absolute abundance for a given date, we adopted a three-step process. First, we used percentiles in the frequency distribution of the cumulated number of reports until the map date (Rodríguez and Delibes 2002) as classes of relative abundance. We assumed a nonlinear relationship between relative and absolute densities for the period 1985–1988 (modified from Rodríguez and Delibes 1990). We derived estimates of lynx absolute abundance (N_1) from classes of relative abundance (Rodríguez and Delibes 1990). Second, we corrected N_1 for the decline in the probability of recording reports with time elapsed since observation. We fitted a generalized linear model of the number of lynx reports per year during the period 1940–1988, using the elapsed time until 1988 as a predictor to estimate, for each year, the expected number of reports we would have obtained had we performed the survey that year. Predicted values were averaged over periods of five consecutive years, and the ratio between predictions for the most recent period (1985–1988) and earlier quinquennia was used as a correction factor. Then we applied the procedure described in step 1, and the corresponding conversion factors, to produce a second estimate of lynx numbers, called N_2. Finally, published figures of lynx abundance were computed by combining estimated densities with precisely outlined distribution patches. To make our coarse-grained estimates (on a grid) comparable to published figures, we multiplied the average proportion of the cell area covered by outlined patches of lynx distribution (0.54) by N_2 to produce the final estimate of lynx numbers per cell. We employed estimates by Simón et al. (2012) for 2002 (approximated here to year 2000), 2005, and 2010, whereas figures for 2015 were published on the website http://www.iberlince.eu/. (last accessed August 25, 2017) For estimates in 2000 or later, total population size refers to the number of adult females, and juveniles plus the number of adult males assuming a 1:1 sex ratio. Isolation date between genetically similar spatial clusters was defined as the date of loss of physical contact in a 10-km UTM grid.

Population-Based Microsatellite Data Analyses

In order to assess structure through time and across space, we used the Bayesian clustering method implemented in STRUCTURE (Pritchard et al. 2000; Falush et al. 2003) to infer the number of genetic clusters in our data set, and to obtain assignment values for each individual in each cluster (*Q*). We also performed FCA as implemented in Genetix (Belkhir et al. 1996–2004) and plotted them with the threejs R package (https://github.com/bwlewis/rthreejs/blob/master/R/threejs.R; last accessed August 25, 2017) to visualize overall patterns of genetic structure in our data. We calculated diversity parameters for each of the geographical groups and temporal periods using Genetix 4.05 (Belkhir et al. 1996–2004) and HP-Rare (Kalinowski 2005). Statistical comparisons in diversity between populations or periods were performed using non-parametric Wilcoxon signed rank tests. In order to evaluate which of the scenarios is the most plausible to explain the genetic differentiation between the populations, pure genetic drift or drift-migration equilibrium, we used a Bayesian–MCMC approach implemented in *2mod* (Ciofi et al. 1999) to estimate the relative likelihood of each model.

Mitogenome Sequence Analyses

Consensus mitogenome sequences were aligned and collapsed to distinct haplotypes using pegas R package (Paradis 2010). Mitogenomic consensus sequences for each sample have been deposited in GeneBank under accession numbers KX911255–KX911412. Diversity and differentiation indices were estimated with PopGenome (Pfeifer et al. 2014) and ape (Paradis et al. 2004) R. Empirical distributions were generated to determine sampling variance and standard deviations. We tested whether haplotype diversity had declined over time using the conservative double-testing implemented in TEST_H_DIFF R (http://www.ucl.ac.uk/tcga/software/; last accessed August 25, 2017). Phylogenetic relationships among sequence haplotypes were inferred by constructing a median-joining haplotype network (Bandelt et al. 1999) using pegas R package (Paradis 2010).

Individual-Based Microsatellite Data Analyses

To assess changes in microsatellite genetic diversity over time, we calculated individuals' standardized observed heterozygosity with IRmacroN4 (Amos et al. 2001) and assessed its relationship with time using linear regression. Changes in allelic frequencies were assessed by testing an isolation-by-time hypothesis, that is, we assessed the correlation between a genetic distance matrix—linearized inter-individual pairwise distance described as $\hat{a}$ by Rousset (2000), as implemented in SPAGeDi (Hardy and Vekemans 2002)—and a temporal distance matrix in years. Significance of the regression slopes was assessed with Mantel tests. We performed this analysis with the four populations for which we had enough samples from a wide enough temporal range (Doñana, Central Range, Montes de Toledo and Eastern Sierra Morena).

For a more detailed version of methods, see supplementary methods, Supplementary Material online.

Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution* online.

Acknowledgments

This research was funded by the Spanish Dirección General de Investigación Científica y Técnica, through projects CGL2006-10853/BOS, CGL2010-21540/BOS, and CGL2013-47755-P of the Spanish Ministerio de Ciencia e Innovación (MICINN) to J.A.G., and by the Fundación BBVA through a project on ancient lynx genetics to M.D. M.C. and E.M. received a JAE predoctoral grant from CSIC (Spanish National Research Council), M.L.P. was supported by PhD contracts from Programa Internacional de Becas "La Caixa-Severo Ochoa." J.N. received financial support through projects HAR2014-55131 from the Ministerio de Ciencia e Innovación (MICINN) and SGR2014-108 from Generalitat de Catalunya. We thank the following people and institutions for contributing specimens information and/or samples: Estación Biológica de Doñana—CSIC, Swedish Natural History Museum, Natural History Museum of London, Smithsonian Natural History Museum, Museo Nacional de Ciencias Naturales—CSIC, Colegio Nuestra Señora de los Reyes, Instituto Aguilar y Eslava, Instituto Padre Luís Coloma, Museo de Ciencias del Instituto Padre Suárez, Facultad de Ciencias de Granada, Museo de Historia Natural de Viso del Marqués, Sociedad Aranzadi, Palacio Real de Riofrío, Museu da Faculdade de Ciências do Porto, Museu Zoológico de Coimbra, Museu Nacional de História Natural (Portugal), Museu Municipal de Penamacor, José Antonio Riquelme, Grupo MIDAS III Milenio, Universidad de Huelva, Ayuntamiento de Úbeda, Instituto Português de Arqueologia, Tapada Nacional de Mafra, Clube Português de Canicultura, Museu Municipal de Moià, Museu do Instituto Geológico e Mineiro (Portugal), Museo Arqueológico Nacional, Prof. Ana Margarida Arruda, Câmara Municipal de Lagos, Ostrava Zoo (Check Republic) and Zoo de Córdoba (Spain). We equally acknowledge all the anonymous owners of lynx specimens who allowed us to sample them; many of these were located with the help of Juan C. del Olmo, Javier N. Guzmán and Margarida Fernandes. We thank Ana Martínez Fontúrbel and Margarida Fernandes for field assistance. Contemporary samples were received under agreements with the Consejería de Medio Ambiente de la Junta de Andalucía, and were contributed by the LIFE projects LIFE-NAT 06/E/000209 and LIFE10 NAT/E/000570, the Iberian lynx Ex Situ Conservation Program, or by Alejandro Rodríguez, Francisco Palomares and José V. López-Bao.

Author Contributions

J.A.G, A.R., E.R., and M.D. designed the study. M.C.-M., E.M. M.L.P., and J.A.G wrote the manuscript. J.A.G, M.S., and M.H. directed the experimental work. M.C.-M., E.M., L.S., and M.L.P. produced the molecular data. M.C.-M., E.M., L.S., B.M.-C., and J.A.G. analyzed the data. A.R. quantified demographic changes. F.N., A.R.-H., A.C., J.N. C.D., E.B.-S., C.F.-R., and M.P.-R. excavated the ancient samples and provided archaeological expertise. All authors contributed to the final version of the manuscript.

References

Abascal F, Corvelo A, Cruz F, Villanueva-Canas JL, Vlasova A, Marcet-Houben M, Martinez-Cruz B, Cheng JY, Prieto P, Quesada V. 2016. Extreme genomic erosion after recurrent demographic bottlenecks in the highly endangered Iberian lynx. *Genome Biol.* 17(1):251.

Allendorf FW, Luikart G. 2007. Conservation and the genetics of populations. Malden (MA): Blackwell.

Allendorf FW, Luikart G, Aitken SN. 2013. Conservation and the genetics of populations. Hoboken: John Wiley & Sons.

Altuna J. 1972. Fauna de Mamíferos de los Yacimiento Prehistórico de Guipúzcoa. *Munibe* 24:1–472.

Amos W, Wilmer JW, Fullard K, Burg TM, Croxall JP, Bloch D, Coulson T. 2001. The influence of parental relatedness on reproductive success. *Proc R Soc B Biol Sci.* 268(1480):2021–2027.

Bandelt HJ, Forster P, Rohl A. 1999. Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol.* 16(1):37–48.

Beaumont MA, Zhang WY, Balding DJ. 2002. Approximate Bayesian computation in population genetics. *Genetics* 162(4):2025–2035.

Belkhir K, Borsa P, Chikhi L, Raufaste N, Bonhomme F. 1996–2004 GENETIX 4.05, logiciel sous Windows TM pour la génétique des populations: Laboratoire Génome, Populations, Interactions, CNRS UMR 5171, Université de Montpellier II, Montpellier (France).

Breitenmoser U. 1998. Large predators in the Alps: the fall and rise of man's competitors. *Biol Conserv.* 83(3):279–289.

Casas-Marce M, Revilla E, Fernandes M, Rodríguez A, Delibes M, Godoy JA. 2012. The value of hidden scientific resources: preserved animal specimens from private collections and small museums. *Bioscience* 62(12):1077–1082.

Casas-Marce M, Revilla E, Godoy JA. 2010. Searching for DNA in museum specimens: a comparison of sources in a mammal species. *Mol Ecol Resour.* 10(3):502–507.

Casas-Marce M, Soriano L, López-Bao JV, Godoy JA. 2013. Genetics at the verge of extinction: insights from the Iberian lynx. *Mol Ecol.* 22(22):5503–5515.

Ceballos G, Ehrlich PR, Barnosky AD, García A, Pringle RM, Palmer TM. 2015. Accelerated modern human-induced species losses: entering the sixth mass extinction. *Sci Adv.* 1(5):e1400253.

Ciofi C, Beaumont MA, Swingland IR, Bruford MW. 1999. Genetic divergence and units for conservation in the Komodo dragon Varanus komodoensis. *Proc R Soc B Biol Sci.* 266(1435):2269–2274.

Clavero M, Delibes M. 2013. Using historical accounts to set conservation baselines: the case of lynx species in Spain. *Biodiv Conserv.* 22(8):1691–1702.

Cornuet JM, Pudlo P, Veyssier J, Dehne-Garcia A, Gautier M, Leblois R, Marin JM, Estoup A. 2014. DIYABC v2.0: a software to make approximate Bayesian computation inferences about population history using single nucleotide polymorphism, DNA sequence and microsatellite data. *Bioinformatics* 30(8):1187–1189.

Drummond AJ, Suchard MA, Xie D, Rambaut A. 2012. Bayesian Phylogenetics with BEAUti and the BEAST 1.7. *Mol Biol Evol.* 29(8):1969–1973.

Ellis EC, Goldewijk KK, Siebert S, Lightman D, Ramankutty N. 2010. Anthropogenic transformation of the biomes, 1700 to 2000. *Glob Ecol Biogeogr* 19:589–606.

Falush D, Stephens M, Pritchard JK. 2007. Inference of population structure using multilocus genotype data: dominant markers and null alleles. *Mol Ecol Notes* 7(4):574–578.

Falush D, Stephens M, Pritchard JK. 2003. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* 164(4):1567–1587.

Fay JC, Wu CI. 1999. A human population bottleneck can account for the discordance between patterns of mitochondrial versus nuclear DNA variation. *Mol Biol Evol.* 16(7):1003–1005.

Ferreras P, Rodríguez A, Palomares F, Delibes M. 2010. Iberian lynx: the uncertain future of a critically endangered cat. In: MacDonald DW, Loveridge AJ, editors. The biology and conservation of wild felids. Oxford: Oxford University Press. p. 507–520.

Fordham DA, Akcakaya R, Brook BW, Rodriguez A, Alves PC, Civantos E, Trivino M, Watts MJ, Araujo MB. 2013. Adapted conservation measures are required to save the Iberian lynx in a changing climate. *Nat Clim Change* 3(10):899–903.

Frankham R, Ballou JD, Briscoe DA. 2010. Introduction to conservation genetics. Cambridge (NY): Cambridge University Press.

Frankham R, Ballou JD, Briscoe DA. 2002. Introduction to conservation genetics. Cambridge (NY): Cambridge University Press.

Frankham R, Ballou JD, Eldridge MDB, Lacy RC, Ralls K, Dudash MR, Fenster CB. 2011. Predicting the probability of outbreeding depression. *Conserv Biol.* 25(3):465–475.

Fu YX, Li WH. 1993. Statistical Tests of Neutrality of Mutations. *Genetics* 133:693–709.

Garrison E, Marth G. 2012. Haplotype-based variant detection from short-read sequencing. In: arXiv.

Godoy JA, Casas-Marcé M, Fernández J. 2009. Genetic issues in the implementation of the Iberian lynx Ex situ Conservation Programme. In: Vargas A, Breitenmoser C, Breitenmoser U, editors. Iberian lynx ex situ conservation: an interdisciplinary approach. Madrid: Fundación Biodiversidad in collaboration with: International Union for Conservation of Nature, Species Survival Commission (IUCN/SSC) Cat Specialist Group. p. 86–99.

Hamilton JA, Miller JM. 2016. Adaptive introgression as a resource for management and genetic conservation in a changing climate. *Conserv Biol.* 30(1):33–41.

Hardy OJ, Vekemans X. 2002. SPAGEDi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. *Mol Ecol Notes* 2(4):618–620.

Hedrick PW. 2001. Conservation genetics: where are we now? *Trends Ecol Evol.* 16(11):629–636.

Hofman CA, Rick TC, Fleischer RC, Maldonado JE. 2015. Conservation archaeogenomics: ancient DNA and biodiversity in the Anthropocene. *Trends Ecol Evol.* 30(9):540–549.

Hofreiter M, Barlow A, Paijmans J, Westbury M. 2015. Genomic analyses from highly degraded DNA. *Genome* 58: 228–228.

Jiménez MA, Sánchez B, Alenza MDP, García P, López JV, Rodríguez A, Muñoz A, Martínez F, Vargas A, Peña L. 2008. Membranous glomerulonephritis in the Iberian lynx (*Lynx pardinus*). *Vet Immunol Immunopathol.* 121(1–2):34–43.

Johnson WE, Godoy JA, Palomares F, Delibes M, Fernandes M, Revilla E, O'Brien SJ. 2004. Phylogenetic and phylogeographic analysis of Iberian lynx populations. *J Hered.* 95(1):19–28.

Jonsson H, Ginolhac A, Schubert M, Johnson PLF, Orlando L. 2013. mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics* 29:1682–1684.

Kalinowski ST. 2005. HP-RARE 1.0: a computer program for performing rarefaction on measures of allelic richness. *Mol Ecol Notes* 5(1):187–189.

Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, et al. 2012. Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28(12):1647–1649.

Leonard JA. 2008. Ancient DNA applications for wildlife conservation. *Mol Ecol.* 17(19):4186–4196.

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 25(14):1754–1760.

Maricic T, Whitten M, Paabo S. 2010. Multiplexed DNA sequence capture of mitochondrial genomes using PCR products. *PLoS ONE* 5(11):e14004.

Martínez F, Manteca X, Pastor J. 2013. Retrospective study of morbidity and mortality of captive Iberian lynx (*Lynx pardinus*) in the *ex situ* conservation programme (2004–June 2010). *J Zoo Wildl Med.* 44(4):845–852.

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, et al. 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20(9):1297–1303.

McLellan R. 2014. Living planet report 2014. In: McLellan R, editor. Living planet report: World Wildlife Fund. p. 178.

Méndez M, Vögeli M, Tella JL, Godoy JA. 2014. Joint effects of population size and isolation on genetic erosion in fragmented populations: finding fragmentation thresholds for management. *Evol Appl.* 7(4):506–518.

Meyer M, Briggs AW, Maricic T, Hober B, Hoffner BH, Krause J, Weihmann A, Paabo S, Hofreiter M. 2008a. From micrograms to picograms: quantitative PCR reduces the material demands of high-throughput sequencing. *Nucleic Acids Res.* 36(1):e5.

Meyer M, Kircher M. 2010. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harbor Protoc.* 2010(6):pdb.prot5448.

Meyer M, Stenzel U, Hofreiter M. 2008b. Parallel tagged sequencing on the 454 platform. *Nat Protoc.* 3(2):267–278.

Nowell K, Jackson P. 1996. Wild cats: status survey and conservation action plan. Gland, Switzerland: IUCN.

Orlando L, Gilbert MTP, Willerslev E. 2015. Applications of next-generation sequencing reconstructing ancient genomes and epigenomes. *Nat Rev Genet.* 16(7):395–408.

Palomares F, Godoy JA, López-Bao JV, Rodríguez A, Roques S, Casas-Marce M, Revilla E, Delibes M. 2012. Possible extinction vortex for a population of Iberian lynx on the verge of extirpation. *Conserv Biol.* 26(4):689–697.

Palomares F, Rodríguez A, Revilla E, López-Bao JV, Calzada J. 2011. Assessment of the conservation efforts to prevent extinction of the *Iberian lynx*. *Conserv Biol.* 25(1):4–8.

Paradis E. 2010. pegas: an R package for population genetics with an integrated-modular approach. *Bioinformatics* 26(3):419–420.

Paradis E, Claude J, Strimmer K. 2004. APE: analyses of phylogenetics and evolution in R language. *Bioinformatics* 20(2):289–290.

Peña L, García P, Jiménez MA, Benito A, Alenza MDP, Sánchez B. 2006. Histopathological and immunohistochemical findings in lymphoid tissues of the endangered Iberian lynx (*Lynx pardinus*). *Compar Immunol Microbiol Infect Dis* 29(2–3):114–126.

Pfeifer B, Wittelsburger U, Ramos-Onsins SE, Lercher MJ. 2014. PopGenome: an efficient Swiss army knife for population genomic analyses in R. *Mol Biol Evol.* 31(7):1929–1936.

Piggott MP, Bellemain E, Taberlet P, Taylor AC. 2004. A multiplex pre-amplification method that significantly improves microsatellite amplification and error rates for faecal DNA in limiting conditions. *Conserv Genet.* 5(3):417–420.

Pimm SL, Jenkins CN, Abell R, Brooks TM, Gittleman JL, Joppa LN, Raven PH, Roberts CM, Sexton JO. 2014. The biodiversity of species and their rates of extinction, distribution, and protection. *Science* 344(6187):1246752.

Pritchard JK, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus genotype data. *Genetics* 155(2): 945–959.

Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26(6):841–842.

Ratkiewicz M, Matosiuk M, Saveljev AP, Sidorovich V, Ozolins J, Mannil P, Balciauskas L, Kojola I, Okarma H, Kowalczyk R, et al. 2014. Long-range gene flow and the effects of climatic and ecological factors on genetic structuring in a large, solitary carnivore: the Eurasian lynx. *PLoS ONE* 9(12):e115160.

Rodríguez A, Calzada J. 2015. *Lynx pardinus*. The IUCN Red List of Threatened Species 2015. doi:e.T12520A50655794.

Rodríguez A, Delibes M. 1992. Current range and status of the Iberian lynx Felis-Pardina Temminck, 1824 in Spain. *Biol Conserv.* 61(3):189–196.

Rodríguez A, Delibes M. 1990. El lince ibérico (Lynx pardina) en España: distribución y problemas de conservación. Madrid, Spain: Instituto Nacional para la Conservación de la Naturaleza.

Rodríguez A, Delibes M. 2002. Internal structure and patterns of contraction in the geographic range of the *Iberian lynx*. *Ecography* 25(3):314–328.

Rodríguez R, Ramírez O, Valdiosera CE, García N, Alda F, Madurell-Malapeira J, Marmi J, Doadrio I, Willerslev E, Gotherstrom A, et al.

2011. 50,000 years of genetic uniformity in the critically endangered *Iberian lynx*. *Mol Ecol.* 20(18): 3785–3795.

Rodríguez-Varela R, Tagliacozzo A, Ureña I, García N, Crégut-Bonnoure E, Mannino MA, Arsuaga JL, Valdiosera C. 2015. Ancient DNA evidence of Iberian lynx palaeoendemism. *Q Sci Rev.* 112:172–180.

Rousset F. (647 co-authors). 2000. Genetic differentiation between individuals. *J. Evol. Biol.* 13:58–62.

Ruiz-López MJ, Gañán N, Godoy JA, Del Olmo A, Garde J, Espeso G, Vargas A, Martínez F, Roldán ERS, Gomendio M. 2012. Heterozygosity-fitness correlations and inbreeding depression in two critically endangered mammals. *Conserv Biol.* 26(6):1121–1129.

Schmidt K, Ratkiewicz M, Konopiński MK. 2011. The importance of genetic variability and population differentiation in the Eurasian lynx *Lynx lynx* for conservation, in the context of habitat and climate change. *Mammal Rev.* 41(2):112–124.

Simón M, Gil-Sánchez J, Ruiz G, Garrote G, McCain E, Fernández L, López-Parra M, Rojas E, Arenas-Rojas R, Rey T, et al. 2012. Reverse of the decline of the endangered *Iberian lynx*. *Conserv Biol* 26(4):731–736.

Tajima F. 1989. Statistical-Method for Testing the Neutral Mutation Hypothesis by DNA Polymorphism. *Genetics* 123:585–595.

Thomas CD, Williamson M. 2012. Extinction and climate change. *Nature* 482(7386):E4–E5.

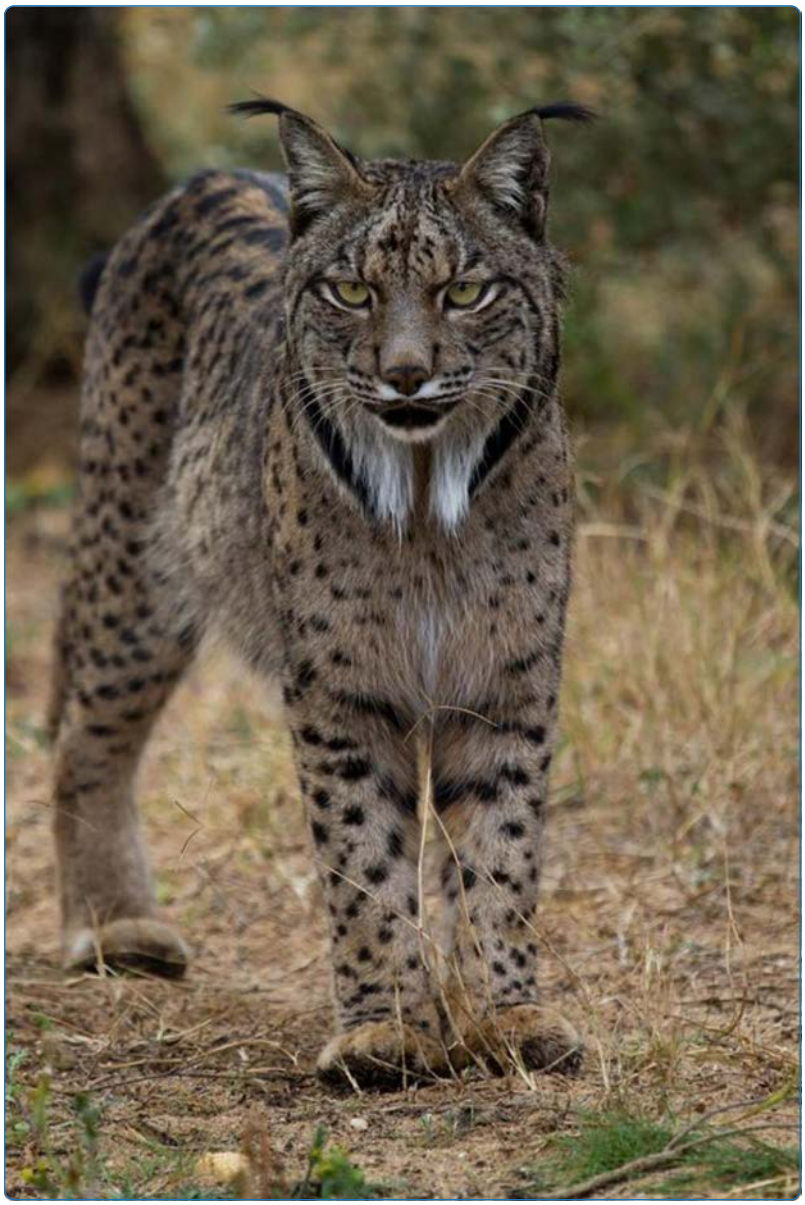
Thomas MA, Roemer GW, Donlan CJ, Dickson BG, Matocq M, Malaney J. 2013. Gene tweaking for conservation. *Nature* 501(7468):485–486.

Valdiosera CE, Garcia-Garitagoitia JL, Garcia N, Doadrio I, Thomas MG, Hanni C, Arsuaga JL, Barnes I, Hofreiter M, Orlando L, et al. 2008. Surprising migration and population size dynamics in ancient Iberian brown bears (*Ursus arctos*). *Proc Natl Acad Sci U S A.* 105(13):5123–5128.

Vigne J-D, Pascal M. 2003. Le Lynx pardelle: *Lynx pardinus* (Temminck, 1827). In: Pascal M, Lorvelec O, Vigne J-D, Keith P, Clergeau P, editors. Évolution holocène de la faune de Vertébrés de France: invasions et disparitions. Paris: Institut National de la Recherche Agronomique, Centre National de la Recherche Scientifique, Muséum National d'Histoire Naturelle. p. 87–88.

Yravedra J. 2005. Aprovechamiento cárnico de lince (*Lynx pardina*) durante el Pleistoceno Superior en el interior de la Península Ibérica. *Munibe* 75:303–311.

Genome Biology



Picture credit: Program Ex-situ Conservation, CC 3.0 Attribution Spain

Extreme genomic erosion after recurrent demographic bottlenecks in the highly endangered Iberian lynx

Abascal *et al.*

BioMed Central

Abascal *et al. Genome Biology* (2016) 17:251
DOI 10.1186/s13059-016-1090-1

Abascal et al. Genome Biology (2016) 17:251
DOI 10.1186/s13059-016-1090-1

Genome Biology

RESEARCH Open Access

Extreme genomic erosion after recurrent demographic bottlenecks in the highly endangered Iberian lynx

Federico Abascal[1†], André Corvelo[2†], Fernando Cruz[2,3,6†], José L. Villanueva-Cañas[4], Anna Vlasova[5,6], Marina Marcet-Houben[5,6], Begoña Martínez-Cruz[3], Jade Yu Cheng[7], Pablo Prieto[5,6], Víctor Quesada[8], Javier Quilez[9], Gang Li[10], Francisca García[11], Miriam Rubio-Camarillo[1], Leonor Frias[2], Paolo Ribeca[2], Salvador Capella-Gutiérrez[5,6], José M. Rodríguez[1,12], Francisco Câmara[5,6], Ernesto Lowy[13], Luca Cozzuto[13], Ionas Erb[5,6], Michael L. Tress[1], Jose L. Rodriguez-Ales[5,6], Jorge Ruiz-Orera[4], Ferran Reverter[5,6], Mireia Casas-Marce[3], Laura Soriano[3], Javier R. Arango[8], Sophia Derdak[2], Beatriz Galán[14], Julie Blanc[2], Marta Gut[2], Belen Lorente-Galdos[9], Marta Andrés-Nieto[15], Carlos López-Otín[8], Alfonso Valencia[1,12], Ivo Gut[2], José L. García[14], Roderic Guigó[5,6,16], William J. Murphy[10], Aurora Ruiz-Herrera[15,17], Tomas Marques-Bonet[2,9,18], Guglielmo Roma[13], Cedric Notredame[5,6], Thomas Mailund[7], M. Mar Albà[4,6,18], Toni Gabaldón[5,6,18], Tyler Alioto[2,6] and José A. Godoy[3*]

Abstract

Background: Genomic studies of endangered species provide insights into their evolution and demographic history, reveal patterns of genomic erosion that might limit their viability, and offer tools for their effective conservation. The Iberian lynx (*Lynx pardinus*) is the most endangered felid and a unique example of a species on the brink of extinction.

Results: We generate the first annotated draft of the Iberian lynx genome and carry out genome-based analyses of lynx demography, evolution, and population genetics. We identify a series of severe population bottlenecks in the history of the Iberian lynx that predate its known demographic decline during the 20th century and have greatly impacted its genome evolution. We observe drastically reduced rates of weak-to-strong substitutions associated with GC-biased gene conversion and increased rates of fixation of transposable elements. We also find multiple signatures of genetic erosion in the two remnant Iberian lynx populations, including a high frequency of potentially deleterious variants and substitutions, as well as the lowest genome-wide genetic diversity reported so far in any species.

Conclusions: The genomic features observed in the Iberian lynx genome may hamper short- and long-term viability through reduced fitness and adaptive potential. The knowledge and resources developed in this study will boost the research on felid evolution and conservation genomics and will benefit the ongoing conservation and management of this emblematic species.

Keywords: Conservation genomics, Genetic diversity, Inbreeding, Genetic drift, Lynx

* Correspondence: godoy@ebd.csic.es
†Equal contributors
[3]Department of Integrative Ecology, Doñana Biological Station (EBD), Spanish National Research Council (CSIC), C/ Americo Vespucio, s/n, 41092 Sevilla, Spain
Full list of author information is available at the end of the article

Background

Species are becoming extinct at rates unprecedented in recent history as a consequence of human activity [1]. Surviving populations of vertebrate species have decreased in size by an average of 58% from 1970 to 2012 [2] and 15–40% of living species are predicted to go extinct by 2050 [3]. While the primary causes of these declines are usually known and have been the main targets of conservation efforts, genetic changes accumulated during the decline can compromise the recovery of endangered populations and limit their long-term viability. Indeed, endangered populations typically show patterns of low genetic diversity and high inbreeding that can result in loss of adaptive potential, reduced rates of reproduction and survival, and increased extinction risk [4]. Genomic approaches are expected to improve our understanding of how the interaction between genetic drift, mutation, recombination, and natural selection shapes the genome of endangered populations and to contribute to a more effective conservation by facilitating the identification and subsequent management of deleterious variants. The fulfillment of these expectations requires, however, genomic studies in well-characterized and actively managed endangered species to serve as models [5].

The Iberian lynx (*Lynx pardinus*) is one of the four extant lynx species that share a short bobbed tail, spotted coat, muscular body, long legs, and characteristic tufted ears and beard-resembling ruffs. The Iberian and Eurasian lynx are sister species and the two extant lynxes in Eurasia, having diverged around 1.1 million years ago (Mya) [6, 7]. In contrast to the large, generalist and widespread Eurasian lynx, the Iberian lynx is smaller and a habitat- and prey-specialist, being restricted to the Mediterranean region in the Iberian Peninsula where they prey almost exclusively on rabbits. Supposed to be once fairly abundant and widely distributed across the Iberian peninsula, a steep decline during the second half of the 20th century left less than 100 lynx (less than 62 mature) distributed in the two isolated populations of Doñana and Andújar (Sierra Morena) in Andalusia, southern Spain, leading to its recognition as the most endangered felid in the world [8] and to its classification as "critically endangered" in the 2002, 2006, and 2008 IUCN red lists. Active conservation in the last 14 years, including in situ management of habitat, prey, and non-natural mortality, captive breeding, translocation, and reintroduction programs, has recovered lynx numbers to over 300 (156 mature) individuals in 2012, leading to its downlisting to "endangered" in the 2015 IUCN red list [9].

Previous studies using microsatellite markers documented low genetic diversity, a high inbreeding rate, and a high genetic differentiation between the two populations [10]. The following evidence suggests that these genetic factors are limiting current reproduction and survival rates (inbreeding depression): (i) an increase in the proportion of abnormal sperm with individual inbreeding [11]; (ii) a recent decrease in litter size and survival in Doñana [12]; (iii) a high incidence of membranous glomerulonephritis and lymphoid depletion [13, 14]; and (iv) several deleterious traits with likely genetic bases segregate at moderate to high frequencies in the captive population, including cryptorchidism and an idiopathic epilepsy [15]. This has prompted the translocation of individuals to reconnect the two remnant populations and their mixing in captivity, which has likely contributed to improved reproductive and survival rates. These circumstances make of the Iberian lynx a good model for the emerging field of conservation genomics [16].

We have sequenced, assembled, and annotated a draft genome of an Iberian male named *Candiles*, and re-sequenced another ten Iberian and one Eurasian lynx genomes. In addition, to obtain gene expression data and to assist gene annotation we have characterized the transcriptome of 11 lynx tissues. We use these resources to analyze the marks left by recurrent demographic bottlenecks on the dynamics of transposable elements (TEs), the rates and patterns of nucleotide substitution, and the efficiency of purifying selection. We characterize the genetic diversity in the two remnant Iberian lynx populations and discuss the interplay between demographic history, GC-biased gene conversion, genetic drift, recombination, and selection in a species on the brink of extinction.

Results and discussion

The Iberian lynx reference genome

We assembled the first draft of the Iberian lynx genome (LYPA 1.0) by combining a fosmid-pool sequencing approach [17] with shotgun sequencing of whole-genome fragment libraries on Illumina and 454 platforms (Additional file 1: Sections 1 and 2). With a contig N50 of 68 kb, our Iberian lynx assembly is more contiguous than other felid genome assemblies, including those of the domestic cat, tiger, and cheetah (Additional file 1: Table S4). However, due to the limited amount of long-range linkage information we were able to produce, we could not achieve as high a scaffold N50 (1.52 Mb). Regardless, the completeness of the gene space as assessed by the Core Eukaryotic Genes Mapping Approach (CEGMA) [18] was 95% (98% including partial genes). We annotated the reference genome with protein-coding genes and other structural and functional genomic features, including TEs and small non-coding RNAs (sncRNAs) and long non-coding RNAs (lncRNAs) (Additional file 1: Section 3). More than 98% of the 21,257 protein coding genes identified were functionally

annotated with InterPro, KEGG, or Blast2GO features (Additional file 1: Section 4). We also performed a focused annotation of the degradome, the complete repertoire of proteases in the organism, with computer-assisted manual methods (Additional file 1: Section 5; Additional file 2: Datasheet S1) and analyzed expression patterns of protein-coding genes (Additional file 1: Section 6) and lncRNAs (Additional file 1: Section 7). A small percentage of these genes (<1%) were expressed in at least one tissue in the Iberian lynx but did not show homology to genes in other mammals or vertebrates (Additional file 1: Section 8; Additional file 2: Datasheet S2). These features, along with other layers of genomic information that include synteny, gene expression, and genomic variation, can be accessed interactively through a dedicated genome browser (http://denovo.cnag.cat/genomes/iberian_lynx; Additional file 1: Section 25).

Evolutionary and demographic history

Phylogenomic analyses confirmed the evolutionary relationships among mammals and felids inferred with smaller datasets [6, 7] and estimated the Eurasian and Iberian lynx divergence as 1.5 Mya (95% credibility interval = 0.9–2.2 Mya; Additional file 1: Section 13). We further investigated the divergence of Iberian and Eurasian lynx using CoalHMM [19, 20] ("Methods"; Additional file 1: Section 9) and their demographic history using PSMC [21] and *∂a∂i* [22] ("Methods"; Additional file 1: Section 10). The former analysis yielded higher support for a divergence model with a period of limited gene flow before complete genetic isolation [20] than for an instantaneous speciation model (isolation with no further gene flow [19] ($\Delta AIC = 4 \times 10^3$; Fig. 1a). We found that the demographic history of the Iberian lynx is punctuated by four bottlenecks (Fig. 1b, c; Additional file 1: Section 10). PSMC inferred a first and most drastic population decline that affected both Iberian and Eurasian lynx 700–100 thousand years ago (kya; Fig. 1b); this contraction may have separated the two species, resulting in the onset of the population structure detected by CoalHMM (312.2 kya; 95% confidence interval 323.1–179.4 kya). Subsequently, the demography of both lynx species followed parallel histories with fluctuations apparently associated with glacial cycles and with Eurasian lynx showing slightly larger population sizes than Iberian. A transient increase in effective population sizes during a period of milder climate (130–60 kya; Riss–Würm interglacial) might have favored a range expansion and the interbreeding of Eurasian and Iberian lynxes, causing the signal of gene flow detected in the divergence analyses ($m = 0.15$ migrants per generation in each direction). The subsequent period of progressive cooling (Würm glaciation) apparently caused a second population contraction that may have re-isolated both lynx species. However, genetic interchange apparently did not cease until recently (2.473 kya; 95% confidence interval 126.8–0 kya; Fig. 1a). Certainly, opportunities for interbreeding did exist in the recent past as the two species co-occurred within the Iberian peninsula in historical times [23] and their ranges probably overlapped during the Holocene in southern France and northern Italy, where they may have hybridized [24]. Post-speciation gene flow is becoming progressively recognized as a frequent phenomenon and it has been fairly common among felids in general and lynxes in particular [6], including current occasional hybridization at sites of range overlap (e.g. *Lynx rufus* and *Lynx canadensis* [25]). Therefore, evidence for post-divergence gene flow is not sufficient by itself to question the species-level taxonomic status of the Iberian lynx; further characterization of the patterns, timing, and outcome of admixture will be needed to assess the level of post-zygotic reproductive isolation between these two lynxes.

The analysis with *∂a∂i* was able to capture a third, more recent decline that reduced the Iberian lynx effective population size to less than 300 individuals around 315 years ago (Fig. 1c). A similarly dated bottleneck was detected previously using microsatellite marker data [10] and has also been suggested for the Eurasian lynx [26]. This period is characterized by human population expansions leading to increased persecution of large carnivores, forest destruction, and expansion of agricultural land across Europe [27]. The fourth and most recent bottleneck occurred during the 20th century. This population contraction, attributed to increased direct prosecution and decreases in main prey caused by two successive viral outbreaks, is probably too recent to be recovered by these methods but is well documented in the literature.

Genome evolution

The Iberian lynx karyotype is identical to that of other felids in terms of diploid number (2n = 38) and G-banding [28] (Additional file 1: Section 11). A finer-scale analysis based on genome alignments identified five potential intra-chromosomal and ten potential inter-chromosomal rearrangements between domestic cat and lynx and up to 37 inversions. Alignments to the dog genome indicated that 20 of the inversions and six of the inter/intra-chromosomal rearrangements occurred in the Iberian lynx lineage. PCR amplification and Sanger sequencing empirically validated 8 of the 15 putative chromosomal rearrangements (five lynx-specific; Additional file 1: Section 12; Additional file 2: Datasheet S3).

We investigated the evolution of lynx genes through several complementary approaches. First, we reconstructed the molecular phylogeny of every lynx protein-

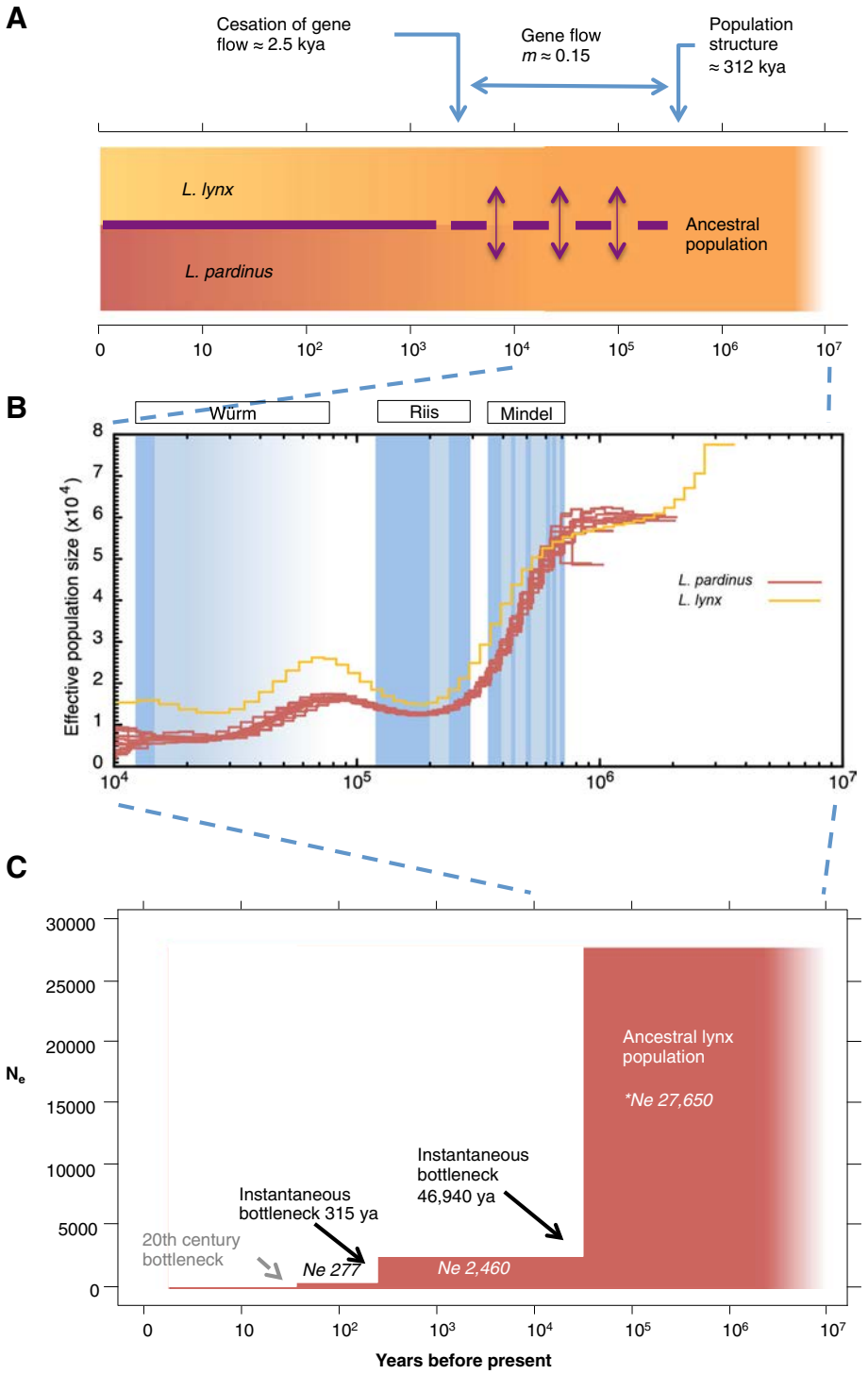


Fig. 1 Evolutionary and demographic history of Iberian and Eurasian lynx. **a** Eurasian and Iberian lynx divergence adjusted to a model of divergence with gene flow. Results suggest a relatively recent divergence of the two lynx species followed by a period of gene flow that lasted until recently (circa 2473 years ago). **b** Effective population size through time estimated for each of the 11 Iberian lynx and a single Eurasian individual using PSMC. Demographic histories are similar for the ten Iberian lynx and slightly different for the Eurasian lynx, although both species follow largely parallel fluctuations of population size probably related to glacial cycles; glacial periods are shaded in blue with glacial maxima in darker tone. **c** Based on the allele frequency spectrum, ∂a∂*i* infers a model with two successive abrupt bottlenecks, one around 47 kya, coincident with the last important decline observed in PSMC output, and a second one at 315 years ago, both reducing to approximately one-tenth the previous population size. The most recent documented bottleneck occurring during the 20th century is not recovered by these methods

coding gene in the context of 15 other mammalian species to establish orthology relationships and to detect and date duplication events [29] (Additional file 1: Section 13). We observed a significant enrichment in genes related to sensory perception of smell among the genes specifically expanded in the lynx lineage and also among those duplicated in the most recent common ancestor of all felids (Additional file 2: Datasheet S4), which is consistent with the importance of smell perception for these carnivores. Besides, and in contrast to humans, felids harbor multiple functional paralogs of the cysteine protease genes *CTSL* and *CTSL2*, ranging from five to ten, which are involved in extracellular matrix homeostasis [30] and immune regulation [31] (Additional file 1: Section 5; Additional file 2: Datasheet S1). Four of these new *CTSL*-like genes have pseudogenized in lynx but not in tiger or cat (Additional file 2: Datasheet S1). Up to 85 additional lynx genes were conservatively identified as putative pseudogenes in lynx (Additional file 2: Datasheet S5).

Second, we used the branch-site test [32] on a set of 9695 one-to-one orthologs from eight mammalian species (*Panthera tigris*, *Felis catus*, *Lynx lynx*, *L. pardinus*, *Ailuropoda melanoleuca*, *Canis lupus familiaris*, *Homo sapiens*, and *Mus musculus*) to identify genes that may have undergone positive selection in lynx (Additional file 1: Section 14). Following extensive manual inspection of the alignments and using strict criteria to minimize alignment errors, we identified 100 genes likely to have accumulated adaptive substitutions in the lynx lineage (Additional file 2: Datasheet S7). Felids possess outstanding hearing [33] and lynx in particular are attributed an exceptionally acute vision and hearing. We found two genes involved in hearing—*CACNA1D* (LYPA23A015140P1) and *MYO1F* (LYPA23A022113P1)—and two genes related to vision—*OPTC* (LYPA23A008195P1) and *GUCY2F* (LYPA23A015393P1) [34]—among those with signatures of positive selection in the lynx lineage.

Population bottlenecks increase inbreeding, reduce diversity, and make purifying selection less effective. Most models predict an increased fixation rate of TEs and a reduced rate of new TE invasions in bottlenecked and inbred populations, although the net effect might depend on the mechanisms of transposition and the relative importance of negative selection and ectopic recombination [35–37]. Our analysis revealed greater expansions of short interspersed elements (SINEs) and long interspersed elements (LINEs) in lynx than in cat and tiger; on the other hand, and in contrast with cat and tiger, we found no clear evidence of recent invasion by new endogenous retroviruses (ERVs) in lynx (Fig. 2a; Additional file 1: Section 15). Our results strongly suggest that the demographic history of the lynx had a strong impact on the TE fixation rate, in accordance with patterns reported for the human lineage [38] and for *Arabidopsis lyrata* [39].

Insertion of TEs within genes is expected to be under stronger purifying selection and more so when it occurs in sense orientation with respect to the gene because sense insertion may disrupt protein translation [40]. Consistent with this and with patterns reported for other genomes [40], LINEs and ERVs, but not SINEs, are particularly depleted within introns in felids and, when present, tend to be in antisense orientation with respect to the gene. Interestingly, lynx-specific LINEs show an increased proportion of insertion within genes in sense orientation (57 out of 112; 51%) compared to the genome-background frequency (37%; $p = 0.004$, Fisher's exact test) (Fig. 2b). A higher proportion of in-sense LINE insertions suggests a less effective purifying selection.

Demographic bottlenecks also reduce heterozygosity and hence are expected to reduce the opportunities for GC-biased gene conversion (gBGC) [41]. The demographic history of Iberian lynx makes it an ideal case to assess the influence of demography on gBGC. We mapped nucleotide substitutions onto the phylogeny of the domestic cat and the Iberian and Eurasian lynx using tiger as outgroup. The weak-to-strong substitution bias (from A/T to G/C; hereafter $W \rightarrow S$), a characteristic signature of gBGC, is generally weaker in the lynx ancestor than in the domestic cat lineage and becomes drastically reduced after the evolutionary split between the Iberian and the Eurasian lynx (Additional file 1: Section 16). This drastic reduction supports a remarkable role for gBGC in shaping the evolution of genomes. Interestingly, the rate of evolution was significantly reduced in both lynx species, suggesting that under population contractions genome stasis may be increased through a decrease in gBGC.

Regions diverging faster (FRs) identified between the cat and the lynx ancestor showed higher $W \rightarrow S$ biases in the faster evolving species (Fig. 3) and were similarly abundant in the two lineages. However, whereas FRs are distributed homogeneously along chromosomal regions in cat, in the lynx ancestor they are concentrated in subtelomeric regions (Additional file 1: Figure S32), as observed in human [42]. In contrast, FRs identified between Iberian and Eurasian lynx showed a lower $W \rightarrow S$ bias in the faster evolving species; they are instead characterized by a reduced heterozygosity (Fig. 2b). We found that these FRs are the result of the differential rate of fixation of ancestral polymorphisms in the two lynx species, which is also supported by an inverse correlation between interspecific ratios of substitution rates and of heterozygosity along the genome ($r = -0.32$, p value $< 2.2 \times 10^{-16}$). The higher number of FRs (117 versus 46) and fixed ancestral

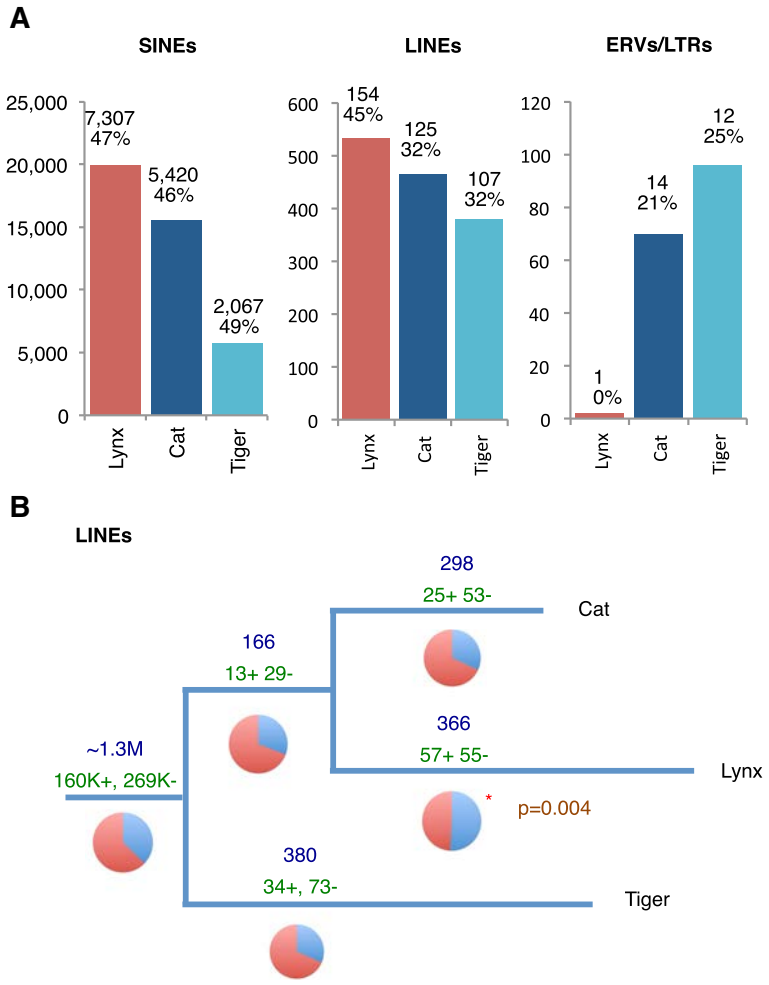


Fig. 2 Comparison of transposable element dynamics. **a** Absolute number of species-specific SINE, LINE, and ERV insertions estimated for lynx, cat, and tiger. For SINEs and LINEs, the number of TE insertions was calculated based on pairwise comparisons between domestic cat and lynx and between domestic cat and tiger. The numbers shown represent the accumulation of insertions in each species since the last common ancestor of lynx, cat, and tiger (Additional file 1: Section 15). The number of TE insertions within genes and the percentage of insertions in sense with respect to the gene are shown on top of each bar. *LTR* long terminal repeat. **b** Phylogenetic tree depicts the number of LINE insertions mapped onto each branch (in *blue above branch*) and the number of insertions that are in sense (+) or antisense (-) orientation with respect to the genes (in *green below branch*). Relative proportions are shown as *pie charts*, with those significantly departing from the average genome-background frequency of 37% (Lynx) indicated by an asterisk. Branch lengths were set manually to reflect the number of LINE insertions on each branch

polymorphisms within FRs (2049 versus 233) in Iberian lynx is consistent with smaller populations sizes and more severe bottlenecks in the Iberian lynx. Seventeen of the Eurasian lynx FRs versus none of the Iberian's were located in subtelomeric regions. Since these 17 FRs did not show high W → S biases, the difference is probably not due to gBGC but to differential loss of heterozygosity in these highly polymorphic regions (see the "Genomic variation" section below).

Ratios of non-synonymous to synonymous substitution rates (dN/dS) are useful means of measuring the strength of purifying selection. Unfortunately, comparison across species is usually not possible because gene annotations differ in their qualities, diminishing the reliability of the alignments. Here, we have developed a new method to select sites aligned with the highest reliability, allowing us to conservatively filter a concatenated alignment of 8117 one-to-one orthologs from different felids, rodents, and hominids (Additional file 1: Section 13). Increased ratios in Iberian ($dN/dS = 0.16$) and Eurasian lynx ($dN/dS = 0.17$) after their divergence from their most recent common ancestor, which has a ratio similar to cat ($dN/dS = 0.06$), are consistent with the relaxation of purifying selection in both species. These ratios are higher than those estimated for other bottlenecked species like humans and chimpanzees (0.10 and 0.11, respectively; Fig. 3c). As most non-synonymous changes are likely to be deleterious, such high dN/dS ratios indicate a high rate of fixation of mildly

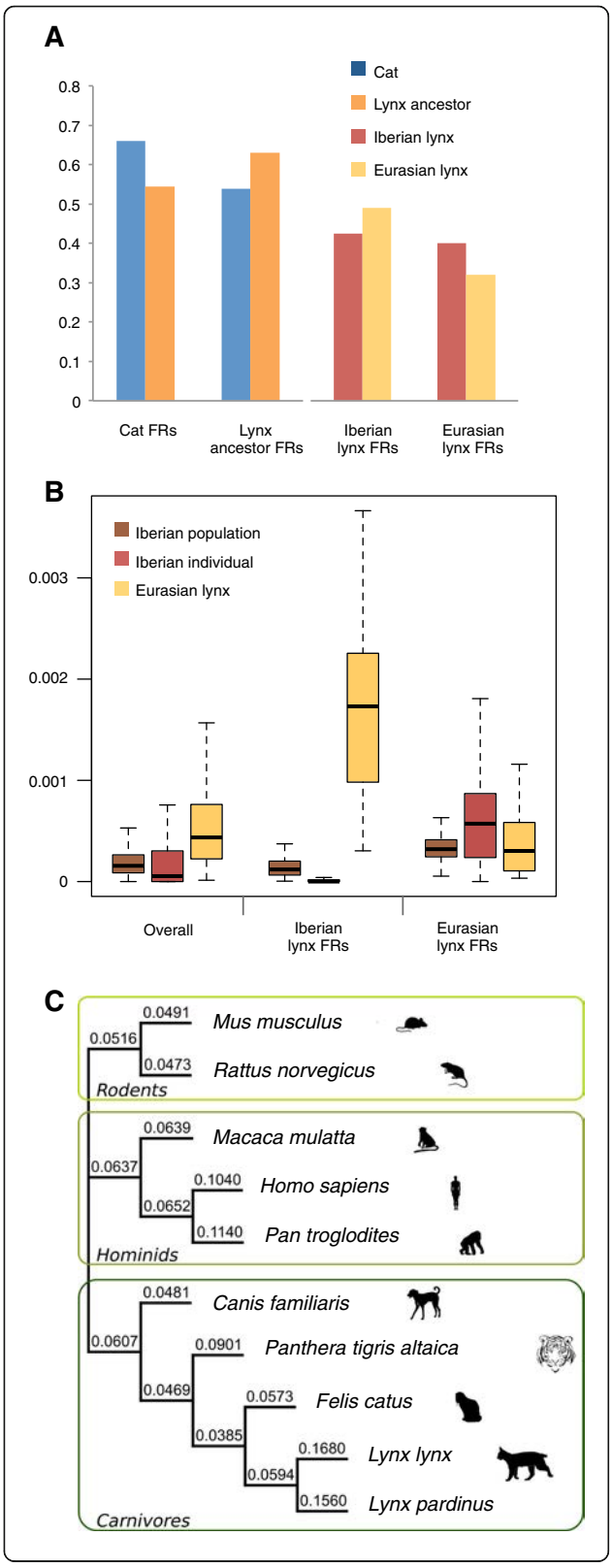


Fig. 3 Patterns of genome evolution. **a** Magnitude of W → S bias within faster evolving regions (*FRs*) identified in pairwise comparisons between cat and the lynx ancestor and between Iberian and Eurasian lynx. **b** Heterozygosity in regions defined as FRs in Eurasian or Iberian lynx. Iberian lynx values are reported for the whole species and for the reference individual. Whereas FRs in cat and the lynx ancestor are associated with higher W → S biases, FRs in both lynx species are associated with reduced heterozygosity and fixation of ancestral polymorphisms. **c** *dN*/*dS* ratios estimated for different mammalian lineages. Increased ratios in lynx indicate the relaxation of purifying selection following the divergence of Iberian and Eurasian lynx

deleterious mutations in both lynx species since their separation from their most recent common ancestor.

Genomic variation

Recent studies have revealed genome-wide signatures of inbreeding, low diversity, and accumulation of potentially deleterious variation in extinct and endangered species, with levels varying extensively, sometimes with little relationship to current demography or conservation status (e.g., [43–47]). To investigate the patterns of genomic variation we identified SNPs using whole-genome shotgun re-sequencing data for 11 Iberian and one Eurasian lynx ("Methods"; Additional file 1: Section 18). Individual Iberian lynx genomes are characterized by the abundance of long runs of homozygosity (ROH; Fig. 4a; Additional file 1: Section 19). The longest ROH (>1 Mb), which are indicative of recent inbreeding, are more abundant in the Doñana population than in the Andújar population (Fig. 4b), resulting in higher average inbreeding coefficients ($F_{ROH\text{-}Doñana} = 0.32$; $F_{ROH\text{-}Andújar} = 0.16$). Medium-length (100 kb–1 Mb) ROH are also more abundant in Doñana, consistent with its lower effective size since the two populations became effectively isolated. Finally, the extent of the genome covered by shorter ROH (10–100 kb) is similar in all individuals, suggesting a shared history of bottlenecks or low population sizes in a more distant past, when the two populations were probably part of a single ancestral population (Fig. 4b).

In line with the analysis of ROH, we found that the linkage disequilibrium (LD), measured as the squared correlation coefficient between genotypes in each individual (r^2), extends to long distances in Iberian lynx (Additional file 1: Appendix, Section 24). r^2 reaches 50% of its maximum value at a distance of 185 kb in Andújar (Fig. 4c), almost twice the average of domestic cat breeds (96 kb) and close to that observed for the highly inbred Burmese cat (249–380 kb) [48]. An even longer extent of LD was estimated for Doñana (1.2 Mb), a result that cannot be solely attributed to its smaller sample size. It must be noted that extensive LD, an additional characteristic signature of small or inbred

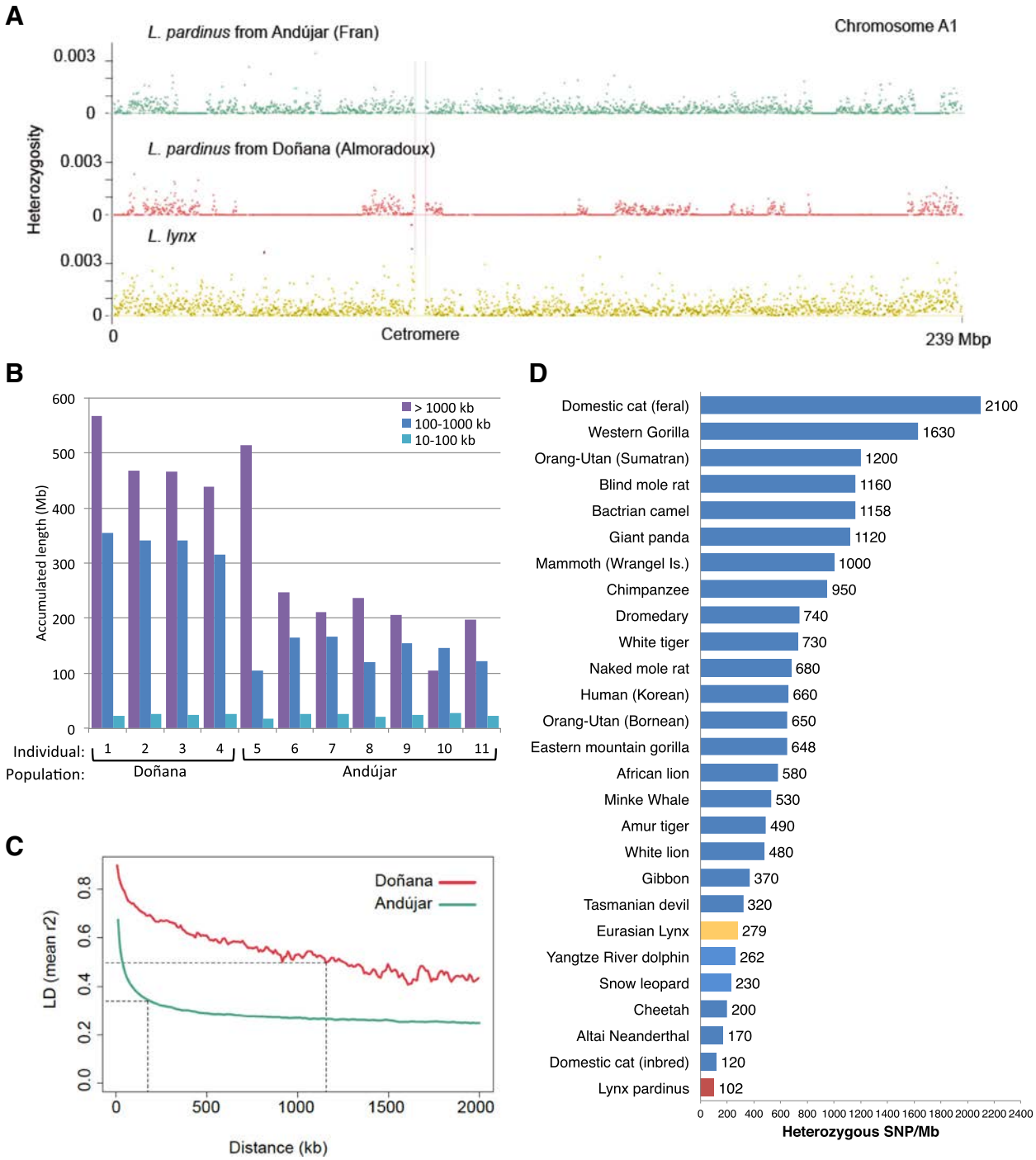


Fig. 4 Patterns of genomic variation in 11 Iberian and one Eurasian lynx. **a** Average heterozygosity in non-overlapping syntenic 100-kb windows in one Iberian lynx from Doñana, one from Andújar (Sierra Morena) and one Eurasian lynx; chromosome A1 is shown as an illustrative example. Long runs of homozygosity are evident in the Iberian individuals. **b** Length of the genome covered by runs of homozygosity of different sizes in each Iberian lynx individual. Both large and medium size ROH are more abundant in Doñana, indicating higher inbreeding and a longer recent history of low effective size. **c** Linkage disequilibrium (*LD*) decay in Iberian lynx populations. Doñana has remained small (50–80 lynxes) and isolated at least since the 1950s, whereas Andújar was part of a large and well connected population until the 1960s; then it became progressively contracted and isolated and reached its lowest size at around 60 animals by 2002. **d** Heterozygous SNP rates in genome-sequenced mammals. Modified from Cho et al. [43] and updated with the addition of data for Altai Neanderthal [114], cheetah and feral domestic cat [115], Yangtze river dolphin [116], gibbon [117], minke whale [118], Eastern mountain gorilla [46], dromedary and Bactrian camel [119], Wrangle Is. mammoth [120], and blind mole rat [121]. The Iberian lynx genome- and species--wide SNP rate and heterozygosity are the lowest reported to date

populations, can hinder the purging of deleterious recessive alleles [49] and may thus have contributed to the accumulation of mildly deleterious mutations in Iberian lynx that we observe.

The two remnant Iberian lynx populations are strongly differentiated ($F_{ST} = 0.22$) and differ in levels of genetic diversity, with Doñana lynxes showing about half the genetic diversity detected in Andújar lynxes (Table 1; Additional file 1: Section 20). Similar patterns were recovered with polymorphic TE insertions (*PhiPT* = 0.261; $H_{E\_Andújar} = 0.240$; $H_{E\_Doñana} = 0.195$; Additional file 1: Section 15) and copy-number variants (Additional file 1: Section 17) and are consistent with previous studies based on microsatellite markers [10]. The average genome-wide heterozygosity rate in the Iberian lynx (102 SNPs/Mb) is the lowest reported for any mammal and is about one-third (36.6%) of that present in the Eurasian lynx (279 SNPs/Mb). Note that this figure is similar to that observed in a highly inbred domestic cat (121 SNPs/Mb) and lower than those of other highly endangered mammals [43] (Fig. 4d) or birds [44], including the endangered crested ibis (*Nipponia nippon*; 430 SNPs/Mb) or the white-tailed eagle (*Haliaeetus albicilla*; 400 SNPs/Mb) [44]. Accordingly, we also observed values of average genome-wide nucleotide diversity and synonymous nucleotide diversity that are to our knowledge the lowest reported for any organism [50] (Table 1; Additional file 1: Sections 20 and 21). At the same time, the ratio of non-synonymous to synonymous diversity is high ($\pi_N/\pi_S = 0.286$), similar to those observed in other bottlenecked populations, such as humans ($\pi_N/\pi_S = 0.241$) [51] or the Galápagos giant tortoise, *Chelonoidis nigra* ($\pi_N/\pi_S = 0.310$) [52], indicating a relative abundance of potentially deleterious mutations segregating at moderate to high frequencies.

To assess whether different parts of the genome might have become differentially affected by genetic drift, we analyzed genetic diversity in non-overlapping 100-kb-long windows along the genome. The X chromosome has been especially depleted of genetic variation: the average X chromosome-to-autosomal normalized diversity ratios at intergenic sites were 0.35 (standard error (SE) = 0.02), 0.29 (SE = 0.02), and 0.38 (SE = 0.03) for the global, Andújar, and Doñana populations, respectively, and ratios were even smaller for coding sequence (Additional file 1: Section 22). Ratios substantially lower than the 0.75 expected at equilibrium are predicted by theory and often observed in recently bottlenecked populations [53].

We also identified regions showing the highest differences in standardized heterozygosity between Eurasian and Iberian lynx ($\Delta Z_H = Z_{H\text{-}Eurasian} - Z_{H\text{-}Iberian}$; Additional file 1: Section 23). Windows within both the 2.5% largest negative (N = 718) and 2.5% largest positive ΔZ_H (N = 671) were significantly depleted of genes (43.2 and 41.1% of outlier windows with genes, respectively, against 50.4% overall; Fisher's exact tests, $p < 0.0001$) and windows with a large negative ΔZ_H ($Z_{H\text{-}Eurasian} << Z_{H\text{-}Iberian}$) were more likely to be in subtelomeric regions (15 versus 6% overall; Fisher's exact test, $p < 0.0001$). A comparatively high diversity in subtelomeric regions supports a role for recombination in maintaining genetic diversity in small or declining populations [54], either through an associated increase in mutation rates or by reducing the number of sites affected by hitch-hiking during positive or purifying selection. Regarding gene functions, a gene ontology enrichment analysis indicated that windows with the largest positive and the largest negative ΔZ_H are both enriched for genes related to olfactory perception and G-protein signal transduction, while windows with the largest negative ΔZ_H ($Z_{H\text{-}Eurasian} << Z_{H\text{-}Iberian}$) were also enriched in genes involved in pheromone reception, amino acid and glucose transmembrane transport, and the regulation of triglyceride biosynthesis genes, among others (Additional file 2: Datasheet S8). The olfactory receptor family is the largest and one of the most genetically diverse multigene families in vertebrates and its evolution has been suggested to be under the influence of balancing selection [55]. Thus, despite the extreme global loss of diversity, functional variation may have been preserved at specific loci (e.g., olfactory receptor genes) by on-going balancing selection, a hypothesis that deserves further investigation.

Table 1 Iberian lynx genetic diversity

	Doñana	Andújar	All
N (chromosomes)	8	14	22
Number of SNPs	625,552	1,383,709	1,587,509
H_O per SNP	0.178	0.317	0.266
Watterson's Θ (%)[a]	0.012	0.022	0.022
H_E per SNP	0.167	0.316	0.336
H_E per site (π) (%)[a]	0.013	0.025	0.026
π_S (%)[b]	0.014	0.026	0.028
π_N/π_S^b	0.287	0.286	0.287

[a]Per site statistics consider the universe of callable sites (2,021,732,768).
[b]Coding sequence estimates are based on 14,028 coding sequences larger than 200 nucleotides.
H_O observed heterozygosity, H_E expected heterozygosity under Hardy–Weinberg equilibrium, π nucleotide diversity, π_S nucleotide diversity at synonymous sites, π_N nucleotide diversity at non-synonymous sites

Conclusions

Our analyses provided novel insights into the evolutionary and demographic history of the Iberian lynx, revealing a recent divergence and continued admixture with the Eurasian lynx and several drastic population bottlenecks in the last millennia. Such demography has shaped the patterns of nucleotide substitution and increased the fixation rate of transposable elements, whereas the predominance of genetic drift and the concomitant decrease in

the efficiency of purifying selection have resulted in extremely low levels of genetic diversity and a high genetic load, indicating a severe level of genomic erosion in Iberian lynx.

The consequences of such low levels of genetic diversity for the viability of the species are hard to predict, but they are likely to limit the capacity of the Iberian lynx to adapt to environmental changes, whereas the excess of deleterious variants in combination with high inbreeding may reduce individual fitness (i.e., inbreeding depression), as suggested by recent evidence [11–15]. Current conservation efforts, including both ex situ and in situ programs, are addressing these threats by promoting the admixture of the two populations and through the genetic management of captive breeding, translocations, and reintroductions. These actions have likely contributed to the recent modest recovery of the population by generating less inbred and more genetically diverse populations that are potentially more fit.

However, these strategies cannot restore the diversity that has been definitively lost, which may limit the species adaptive potential to environmental change. Increasing the adaptability of the species would demand the careful consideration of novel but controversial genetic restoration approaches such as facilitated adaptation, genome editing, or assisted adaptive introgression [56, 57]. Although the evidence for recent natural introgression could encourage the use of Eurasian lynx as a source for the latter, such drastic measure should require the careful evaluation of hybrid fitness and the associated risks of maladaptation and hybrid swarm [58]. In any case, existing examples of species with long-term persistence and widespread distribution despite depleted genetic diversity [59] allow for some measure of hope and argue for the maintenance of current conservation efforts.

The Iberian lynx draft genome, along with the other resources generated in this study, will support the species conservation by providing more informative and efficient genetic markers for genetic monitoring and management, which is currently based on 36 microsatellite markers. A selected set of highly informative SNPs will provide more reliable and cost-effective tools for genetic monitoring from non-invasive samples (e.g., scats or hairs) and more accurate estimates of relatedness and inbreeding for an integral genetic management of the species that should cover ongoing captive breeding, translocations, and reintroductions. Most importantly, these resources will facilitate the identification and eventual management of deleterious alleles of highest impact on Iberian lynx reproduction or survival.

Methods

Samples

Eleven Iberian and one Eurasian lynx were sequenced in this project (Additional file 1: Table S1). A moderately inbred male born in Andújar in 2006 and kept as a founder of the captive population since then (Candiles) was selected to provide the reference genome for the species. Ten additional Iberian and one Eurasian male were sampled for whole-genome resequencing, four of them from the population in Doñana and six from Andújar. These two populations differ in their recent demography: Doñana has remained small and isolated at least since the 1950s, while Andújar is the result of the progressive contraction of the large and more connected population of Sierra Morena [60]. The Eurasian lynx is a male born in captivity in 2007 at the Zoológico de Córdoba (Spain) with no recent history of close inbreeding. Samples for DNA sequencing were obtained from blood and DNA was extracted following standard protocols. Ten organs (brain, heart, kidney, liver, lung, muscle, pancreas, spleen, stomach, and testes) were sampled for RNA sequencing from one of the Doñana Iberian lynx (Almoradux) immediately after its euthanization. Organ samples were immediately frozen in liquid nitrogen and kept at −80 °C. Total RNA was extracted by the RiboPure™ RNA Purification Kit (Ambion®).

Genome sequencing and assembly

Genomic DNA from a single captive male lynx (*Candiles*; studbook no. 0029) was isolated and shotgun-sequenced using Illumina and 454 technology in libraries of insert sizes 500 bp, 4.5 kb, and 5.2 kb. At the time, larger-insert mate-pair libraries, while desirable, were difficult to obtain and we opted for a fosmid-pool strategy [17]. Ninety fosmid pools of 1200 clones prepared using the NxSeq 40 kb Mate-Pair Cloning Kit (Lucigen Corporation, USA) were used for fosmid-end and fosmid-pool sequencing. The 115,000 clones represented an approximately 1.6-fold physical coverage of the genome. Each pool was shotgun sequenced to greater than 100× depth and assembled independently. The resulting contigs were then merged to obtain an assembly representing approximately 67% of the estimated size of the genome. The remaining portion of the genome was assembled using whole-genome shotgun (WGS) paired-end (PE) data. Both partial assemblies were combined by scaffolding with WGS PE and mate-pair data, followed by extra scaffolding steps using RNA-seq and fosmid end data. The CEGMA pipeline was used to determine the state of the gene space as an indicator of genome completeness [18] (Additional file 1: Sections 1 and 2).

Transcriptome analyses

Total RNA was extracted by the RiboPure™ RNA Purification Kit (Ambion®) from ten organs (brain, heart, kidney, liver, lung, muscle, pancreas, spleen, stomach, and testes) sampled immediately after the euthanasia of one Iberian lynx (*Almoradux*) and from the blood of *Candiles*. Sequence libraries were prepared using the

mRNA-Seq sample preparation kit (Illumina Inc., catalog number RS-100-0801). Reads were aligned to the reference assembly using GEMTools RNAseq pipeline v.1.6.2. Flux Capacitor v.1.2.4 [61] was used to quantify genes, transcripts, exons, and splice junctions in each sample separately. Expression levels were obtained in pure read counts and in reads per kilobase per million mapped reads (RPKM) [62]. Differential gene expression (DGE) analysis across tissues was performed with Bioconductor package edgeR v.3.4.2 (R v.3.0.2) [63] using classic pairwise comparison. A comparative gene expression analyses was performed using data from Brawand et al. [64] and RNA-seq of testis transcriptome from the domesticated cat, *F. catus* (Sequence Read Archive experiment ID SRX193575) using NOISeq Bioconductor package v.2.6.0 [65] (Additional file 1: Section 6).

Genome annotation

Transposable elements and other repeats were annotated with RepeatMasker (version open-4.0.1) [66], using rmblastn v2.2.27 search engine and RM database v20120418, with the *F. catus* library of repeats and the sensitive search option. Low-complexity regions were identified with DustMasker v.2.2.28 [67] with default parameters. Protein-coding genes were then annotated by combining transcriptome evidence with homology-based and ab initio gene prediction methods (Additional file 1: Section 3). Ab initio gene predictions were performed on the TE-masked assembly using *Genid*, *SGP2*, *GlimmerHMM*, and *Augustus*. A combination of the Program to Assemble Spliced Alignments (PASA r2012-06-25) and Evidence Modeler (EVM r2012-06-25) [68] was used to obtain consensus coding sequence (CDS) models using three main sources of evidence: aligned transcripts, aligned proteins, and gene predictions. Small structured non-coding RNAs were detected using the CMsearch tool from the Infernal package (version 1.1rc2) [69] against the Rfam database (version 11) [70]. Long non-coding RNAs (lncRNAs) were predicted by homology using the strategy reported in [71, 72] and by ab initio approaches using *Geneid* to generate a final set consisting of transcripts that are either expressed or conserved in at least one species (Additional file 1: Section 7).

Functional annotation

We used our own automatic functional annotation pipeline based on Interproscan [73], KEGG [74], and Reactome [75] and Blast2GO [76] to assign a description (e.g., the protein name) and relevant annotation through sequence similarity and Gene Ontology-based data mining (Additional file 1: Section 4). SignalP [77] was used to predict the presence and location of signal peptide cleavage sites. Finally, in order to organize, store, and facilitate the access to the entire set of annotations we have developed a MySQL (http://www.mysql.com/) relational database. The modules implemented in APPRIS (http://appris.bioinfo.cnio.es/docs/appris.html) were used to map a range of conserved protein features to the splice variants annotated for each gene and to determine which of these is the main (principal) gene variant. The number of genes annotated with protein features is similar to that of the human genome, but less Iberian lynx genes align full length and without gaps to orthologs in other species or contain signal sequences. The main protein isoform could be identified for the majority of lynx genes with multiple variants (3408 of 5218 genes, 64.9%) and 8066 variants were tagged as alternative. A computer-assisted manual annotation of the degradome [78], the complete repertoire of proteases in the organism, found almost all of the 635 expected proteases, of which 306 were completely annotated, confirming a good gene coverage of the Iberian lynx genome (Additional file 1: Section 5).

Orphan genes

We developed a pipeline to identify lynx orphan protein-coding genes (Additional file 1: Section 8). First, we discarded any proteins that had homologs in any of 23 non-mammalian eukaryotic species, using gene protein coding annotations from Ensembl. To search for homologs we used BlastP (2.2.23+) [79] with an E-value threshold of 10^{-4} and the filter for low complexity regions activated. Second, we discarded any proteins for which we could indirectly trace homology to other species through a second protein in lynx. This could happen, for example, if the protein had evolved very rapidly after a gene duplication event [80]. For these searches we used BlastP with the same parameters as previously except that we used a BLOSUM80 matrix instead of the default BLOSUM62, as we were searching for sequences that had diverged relatively recently. Third, we classified the remaining proteins as lynx-specific or mammalian-specific depending on the presence of homologs in the annotated genes from *F. catus*, *C. lupus familiaris*, *A. melanoleuca*, *Mustela putorius furo*, *H. sapiens*, *M. musculus*, *Bos taurus*, *Equus ferus caballus*, and *Myotis lucifugus* (Ensembl version 72). Fourth, we only selected those genes expressed in at least one tissue using a RPKM threshold of 0.3. This resulted in the identification of 323 lynx-specific genes. The current gene catalogs are likely to be incomplete and this means that some of these 323 putatively lynx-specific genes may correspond to not yet annotated genes in other mammals. We thus employed published RNA-seq data for different tissues and mammalian species [64] to have a more comprehensive set of transcripts to compare our genes with. We ran TopHat2 v.2.0.8 [81] for pooled-

tissue reads from human, mouse, chimpanzee, macaque, and orangutan. Next, all long expressed transcripts (length >200 nucleotides) were assembled using Cufflinks (v.2.0.2) [81] for each species and tissue separately, not using information from gene annotations (no reference GTF file). We used Cuffmerge to obtain a comprehensive set of transcripts for each species and Cuffcompare to classify the transcripts into already known transcripts (annotated, using GTF files corresponding to Ensembl v.60) and novel transcripts (non-annotated). Finally, we ran tBlastX with an E-value threshold of 10^{-6} to search for homologs of the 323 putative lynx orphan genes among these transcripts. After discarding any gene that had at least one match, the list of lynx orphan genes was reduced to 204 (206 transcripts).

Evolutionary and demographic history

We extracted the autosomal contigs and obtained maximum likelihood estimates for either the isolation model of Mailund et al. [19] or the initial migration model of Mailund et al. [20] (Additional file 1: Section 9). As in Mailund et al. [20] we used Akaike's information criterion (AIC) to determine the most probable model. To estimate parameters we used a Nelder-Mead optimization as implemented in scipy's optimization module. The scripts used were "isolation-model.py" and "initial-migration-model.py" from https://github.com/mailund/IMCoalHMM. We split the autosomal contigs into 44 sets each covering ~100 Mbp and estimated the uncertainty in the parameter estimates using a leave-one-out jackknife approach. We used two complementary approaches to infer the demographic history of Iberian lynx (Additional file 1: Section 10). The first uses a pairwise sequentially Markov coalescent (PSMC) model applied to complete diploid genome sequences of single individuals to reconstruct the demographic history of the species from the distribution of the local density of heterozygous sites [21]. The method seems to work well for periods between 10,000 to 1 million years b.p. but tends to overestimate recent population sizes and to spread sudden changes in population size over several preceding tens of thousands of years. For the second approach we used the maximum likelihood inference method implemented in the software *∂a∂i*, which searches for the most recent demographic history that better fits the observed allelic frequency spectrum (AFS) [22]. In *∂a∂i* we evaluated either a single or two demographic changes, allowing them to be instantaneous or exponential and chose the best-fit model using the AIC.

Karyotype

Cells from a primary Iberian lynx fibroblast cell line were harvested at early passages and chromosomal preparations were obtained following standard protocols. Metaphases were stained homogenously with Leishman solution for the analysis of diploid number (2n) and the number of autosomal chromosome arms (NFa) and then G-banded with Wright's stain following the methods described by [82] for karyotyping. For each staining, at least 30 metaphase spreads were analyzed. The karyotype of the Iberian lynx was arranged following the cat chromosomal nomenclature [83] and compared to previously published karyotypes for domestic cat and Eurasian lynx [28, 84]. For telomere detection fluorescence in situ hybridization (FISH) analysis was performed using a peptide nucleic acid (PNA) probe complementary to the telomere G-rich strand (TelC; Panagene, Yuseong-gu, Daejeon, Korea) according to the manufacturer's protocol (Additional file 1: Section 11).

Synteny

We built several pairwise alignments between lynx, cat, tiger, and dog genomes using LAST v.458 [85]. For each scaffold, we sorted all the best-hit alignments of length >1000 bp based on the corresponding cat genome coordinates. Then, all the alignments that were less than 20 kb apart and lay on the same strand were merged with bedtools [86]. We retained only those chained alignments that were at least 15 kbp long and for which at least 40% of the sites in the region were aligned. Finally, we explored the resulting chained alignments to detect inversions and inter/intra-chromosomal rearrangements. We used the dog genome as outgroup to determine whether the potential rearrangements took place in the cat or in the lynx branches. To filter rearrangements that may be assembly artifacts we required that at least one scaffold derived from fosmid sequencing crossed the predicted breakpoint. We tested the scaffold integrity by performing long-range PCRs with primers flanking the inferred breakpoint on the reference genome, followed by Sanger sequencing. Out of 15 potential rearrangements tested, eight were empirically validated by this approach (Additional file 2: Datasheet S3; Additional file 1: Section 12).

Phylogenomics

The Iberian lynx phylome (i.e., the complete collection of phylogenetic trees for each gene encoded in the genome) was reconstructed using the PhylomeDB pipeline [87]. We used 15 mammalian species for expansion and pseudogene analyses and 17 for *dN*/*dS* estimation (Additional file 1: Section 13). Maximum likelihood (ML) trees were reconstructed based on the codon alignments using codonPhyML v.1.0 [88] with GY as codon substitution model and F3X4 as model for defining the codon frequency from the alignment. The resulting phylome was used to infer orthology and paralogy relationships. For each tree, ETE v.2 [89] was used to identify duplication and speciation nodes along the trees using a

species overlap approach and a species overlap score of 0, as described by Huerta-Cepas et al. [90]. All orthology and paralogy relationships are available through PhylomeDB [91]. Gene Ontology enrichment analysis was performed using FatiGO [92]. To find putative pseudogenized genes in lynx, domestic cat genes showing no homologs in the Iberian lynx genome, even when relaxing the overlap threshold to 20%, and that had homologs in at least four additional species were searched against the lynx genome using tBlastn [79]. Cat proteins with significant (e-value $<10^{-5}$) hits in the lynx genome aligning over 30% of their length were selected for further inspection. The genomic region determined by the blast search was extended by 10,000 nucleotides at both sides and exonerate-based gene prediction [93] was performed on the region using the cat protein as a seed; 85 predictions interrupted by stop codons were considered as putative pseudogenes.

We used 8117 sets of one-to-one orthologs comprising proteins from five carnivore species (*L. pardinus*, *L. lynx*, *F. catus*, *P. tigris*, and *C. familiaris*), three primates (*H. sapiens*, *P. troglodytes*, and *M. mulatta*), and two rodents (*M. musculus* and *R. norvegicus*) to estimate *dN*/*dS* ratios for different branches of the extended reference species tree (Fig. 3c, main text). To reduce the impact of alignment errors, the trimmed alignments used to reconstruct single gene trees were further filtered using an automated script (selective_trimming_for_dNdS_analyses.based_neighbours.py) available at the official trimAl repository in GitHub (https://github.com/scapella/trimal). Firstly, codon columns containing gaps were removed. Secondly, we scanned the corresponding translated alignments looking for columns with at least one amino acid replacement and only those surrounded by two previous and two posterior fully conserved sites were retained. Resulting alignments were concatenated and the number of nonsynonymous substitutions per nonsynonymous site (*dN*), synonymous substitutions per synonymous site (*dS*), and the corresponding *dN*/*dS* ratio (ω) were estimated for each branch using the ML method implemented in the CodeML program of PAML v.4.4. [32]. For this analysis, we used a (1) fixed topology according to the extended species tree, (2) F3X4 as model of codon frequency, and (3) a free-omega model (model = 1) so an independent ratio for each branch is assumed.

We also produced a dated Felidae phylogeny based on filtered whole-genome alignments of available felid genomes (domestic cat, Iberian lynx, Eurasian lynx, tiger, lion, snow leopard, and cheetah; Additional file 1: Section 13.8). We analyzed the alignment using *Saguaro* to identify chromosomal regions with discrete phylogenetic signals that are different from the background signal [94] (Additional file 2: Datasheet 6), constructed a maximum likelihood tree in RAxML [95], and used the topology matching the species tree, which was also the most frequent one, to estimate divergence times from whole-genome alignment with *MCMCtree* [32].

Positive selection

We looked for signatures of positive selection in the lynx lineage using a set of 9695 one-to-one orthologs generated in the phylogenomics analyses (Additional file 1: Section 14). We selected eight different species: *P. tigris*, *F. catus*, *L. lynx*, *L. pardinus*, *A. melanoleuca*, *C. lupus familiaris*, *H. sapiens*, and *M. musculus*. We performed multiple sequence alignments with the software PRANK [96] and conducted a branch-site test of positive selection (PS) [32] using information from Timetree (http://www.timetree.org/) for the input tree. We filtered out cases with more than one site with a probability of being under positive selection higher than 0.99 by the Bayes empirical Bayes (BEB) approach as they typically corresponded to non-homologous stretches [97]. We manually validated 100 lynx positive selection candidates (96 for *Lynx* sp. and four for *L. lynx*; Additional file 2: Datasheet S7). We used Gitools [98] and annotations from Ensembl version 73 [99] to perform an enrichment analysis in the set of positively selected genes.

Transposable elements

Genomes of lynx, cat, and tiger were pairwise aligned using LAST [85] with the aim of identifying orthologous regions between them (Additional file 1: Section 15). We analyzed unambiguously aligned regions for each pair of species (lynx–cat, lynx–tiger, cat–tiger) to identify strongly supported gaps. Every gap in which a particular TE covered at least 95% of the gap, 99% of that TE was within the gap, and in which target-site duplications (TSDs) were detected at each gap boundary, was considered as a species-specific TE insertion. TSD were defined by obtaining −25/+15 and −15/+25 bp around the start and end site coordinates of the TE, respectively. Then, both sequences were compared to each other with BLAST and we required that L*P/100 was greater than 6, where L is the length of the alignment and P the percentage of identity. This procedure allowed the identification of short interspersed element (SINE) and long interspersed element (LINE) insertions, as they leave clear TSDs of size ~20 bp. To analyze the accumulation of TEs along the branches of the tree that relates lynx, cat, and tiger, we relied on the pairwise comparisons between lynx and cat and between tiger and cat. Every TE insertion was mapped onto the domestic cat genome to analyze the patterns of insertion within genes.

To determine the activity of endogenous retroviruses (ERVs) in lynx, we relied on a combined approach based on synteny analyses and phylogenetic reconstruction.

First, we annotated the set of endogenous retroviruses in lynx, cat, and tiger. To reconstruct full ERVs we post-processed RepeatMasker results and searched for pairs of long-terminal repeats (LTRs) that (1) were of the same type and (2) were on the same strand and (3) for which at least 50% of the LTR-enclosed sequence corresponded to ERV fragments of the same family and orientation. Finally, when ERV candidates overlapped, we retained only one of them. By doing this, we were able to reconstruct 1776, 1895, and 1940 full-ERV candidates in lynx, cat, and tiger, respectively. We built a phylogenetic tree for all these ERVs using the BioNJ method [100] (Additional file 1: Section 15).

Substitution patterns

To identify and polarize substitutions in Eurasian and Iberian lynxes we called variants with the RubioSeq pipeline [101] using the genome of domestic cat (version 6.2, felcat5) as reference (Additional file 1: Section 16). Based on the genotype of each lynx species, we selected all those sites, either variant or invariant with respect to cat, which were reliably predicted in both species in homozygosis (heterozygous sites were treated separately). The resulting dataset encompassed 2.15 billion genotyped base pairs. To infer ancestral character states, we focused on the set of sites lying on regions for which orthology was successfully established between lynx, cat, and tiger; we excluded sites lying on repeats and/or low-complexity regions. The final dataset contained 1,062,208,795 genotyped sites and included: 712,201 and 707,025 variants specific for Iberian (*L. pardinus*) and Eurasian (*L. lynx*) lynx, respectively; 9,687,075 variants shared by the two (substitutions occurring since the divergence of cat and lynxes until the divergence of Iberian and Eurasian lynxes); and 1,051,102,494 shared invariant sites. Identified substitutions were used to estimate substitution rates, non-synonymous to synonymous substitution ratios (*dN*/*dS*), and weak-to-strong (mutations from A/T to G/C; hereafter W → S) substitution biases. We translated cat genome coordinates to lynx scaffold coordinates (based on the genome alignments) to annotate the effect of substitutions on lynx protein-coding genes using SnpEff v.3.5 [102] and based on the principal transcript isoforms identified with APPRIS [103]. Substitutions were condensed into non-overlapping 100-kbp windows containing at least 10,000 informative sites to analyze the patterns of evolution along chromosomes.

Segmental duplications

We detected segmental duplications (SD) in the genomes of one Eurasian lynx (*L. lynx*) and 11 Iberian lynxes (*L. pardinus*) both from Sierra Morena (7) and from Doñana (4) (Additional file 1: Section 17). Illumina 100-bp reads were mapped to the repeat-masked Fca-6.2 (UCSC felCat5) domestic cat assembly using BWA [104] (using as parameters "bwa -q 15") and duplicated reads were removed with SAMtools [105]. Successfully mapped reads in the resulting BAM files were then used to recover the original FASTQ files using the bam2fastq tool (http://gsl.hudsonalpha.org/information/software/bam2fastq). The final set of 100-bp reads were clipped to 36-bp fragments but only retaining positions in the read with high quality, which was assessed with fastqc (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). The resulting 36-bp reads were then mapped to the reference assembly using mrFast [106] (using as mapping parameters "–e 2"). mrCaNaVaR (v.3.0.1) [107] was used to estimate the copy number along the genome from the mapping read depth. Mean read depth per base pair was calculated in 1-kbp non-overlapping windows and a control read depth distribution was obtained by iteratively excluding windows with extreme read depth values relative to the normal distribution. The mean read depth in these control regions was considered to correspond to copy number equal to two and was used to convert the read depth value in each window into a GC-corrected absolute copy number in each sample. We called SDs in each individual as genomic regions in which the predicted copy number significantly exceeded diploidy, while accounting for the technical variation in the copy number predictions across samples. Finally, we filtered out SDs shorter than 10 kbp and with >85% of their size overlapping with repeats.

Variant calling

To generate variation data for population genomics and species divergence analyses we re-sequenced ten Iberian (mean depth 26.4×) and one Eurasian lynx (64×) (Additional file 1: Section 18). Variant discovery and genotyping were performed using a mapping-based approach as implemented in the RUbioSeq suite [101], using either the Iberian lynx or the domestic cat genome as references and different sets of samples. In addition, we applied the reference-free and assembly-based strategy implemented in *Cortex_var* [108]. The two procedures yielded largely concordant results but, as expected by its higher sensitivity to detect singletons, the mapping-based calling dataset yielded more variants, slightly higher diversity estimates, and more reliable estimates of LD and homozygosity blocks and was the one used for population genomic analyses.

Population genomics

We used VCFTools v.0.1.10 [109] to identify runs of homozygosity (ROH) in each Iberian lynx individual, to estimate individual inbreeding coefficients from the extent of ROH larger than 1 Mbp (F_{roh}) and from the

observed homozygosities and allele frequencies (F_h) (Additional file 1: Section 19), and to compute diversity and differentiation parameters (Additional file 1: Section 20). Estimates were averaged for all variants and converted to per-site averages using the number of reliably called invariant sites. For coding sequences we counted synonymous and non-synonymous variants as annotated by SnpEff v.3.5 [102] (P_s, P_n) and calculated the per-site synonymous and non-synonymous nucleotide diversity (π_S, π_N) by assuming that three-quarters of all of the sites are non-synonymous. The genomic averages of π_S and π_N were calculated by averaging across CDS with more than 200 callable sites weighted by the number of sites; genomic ratios were calculated from these averages (Additional file 1: Section 21). To study the distribution of diversity along and across chromosomes, we considered genomic regions with conserved synteny to the domestic cat genome (felCat 6.2) so that we could assign lynx regions to specific chromosomal locations and obtain estimates of cat–lynx divergence (Additional file 1: Sections 22 and 23). We excluded repeats, low complexity regions, centromeres, and telomeres along with 2 Mb of flanking regions, and the pseudoautosomal region 1 (PAR1) in the X chromosome (first 6 Mb), and defined non-overlapping 100-kb-long syntenic windows. For each of these windows we estimated the nucleotide diversity (π, nucleotide diversity) and the divergence to cat (D, the observed fraction of fixed differences), and the ratio between the two (π/D) was used as a measure of diversity normalized by mutation rate. Standard errors were calculated by bootstrapping over windows or CDS as implemented in the *boot* package for R [110, 111] to account for the correlation among nearby sites due to LD (Additional file 1: Section 22).

To characterize genomic patterns of diversity in the Iberian lynx genome in comparison to the Eurasian lynx genome, for each window we calculated the Z-transformed per-site average of the observed heterozygosity in Eurasian ($Z_{H\text{-}EL}$) and in Iberian lynx ($Z_{H\text{-}IL}$) and its difference ($\Delta Z_H = Z_{H\text{-}EL} - Z_{H\text{-}IL}$). Windows with a ΔZ_H value higher than the 99.9th or equal to or lower than the 0.1th percentile were identified as outliers. We then tested whether outlier windows were preferentially located in subtelomeric regions (within 5 Mb from the end of chromosomes) or were significantly enriched for particular cellular components, biological processes, or molecular functions by performing a Gene Ontology analyses using FatiGO [112], as implemented in Babelomics 4.3 [113].

Additional files

Additional file 1: Supplemental information. Additional details on methods and results, including additional tables (Tables S1–S37) and figures (Figures S1–S45). **Section 1:** Samples, libraries and sequencing. **Section 2**: Genome assembly. **Section 3:** Genome annotation. **Section 4:** Functional annotation. **Section 5:** Manual annotation and comparative analysis of lynx protease genes. **Section 6:** Transcriptome characterization. **Section 7:** Evolutionary profiling and expression of lncRNAs. **Section 8**: Lynx orphan genes. **Section 9:** Eurasian and Iberian lynx divergence **Section 10:** Demographic history. **Section 11:** Karyotype. **Section 12:** Genome alignments and synteny analysis. **Section 13:** Phylogenomics. **Section 14:** Positive selection. **Section 15:** Transposable elements dynamics. **Section 16:** Substitution patterns. **Section 17:** Segmental duplications. **Section 18:** Variant discovery and genotype calling. **Section 19:** Runs of homozygosity (ROH) and individual inbreeding. **Section 20:** Genomic averages of population genetics parameters. **Section 21:** Variation and divergence at coding sequences. **Section 22:** X chromosome versus autosomes genetic diversity. **Section 23:** Patterns of diversity across the genome. **Section 24:** Linkage disequilibrium. **Section 25:** The Iberian lynx genome browser. **Section 26:** References. (PDF 10 MB)

Additional file 2: Supplemental datasheets. Datasheet S1: Summary of the annotation of the human and felid degradomes. **Datasheet S2:** List of orphan genes with expression levels in different organs. **Datasheet S3:** List of potential chromosomal rearrangements with respect to domestic cat. **Datasheet S4:** Gene Ontology terms enriched in proteins that duplicated at specific points in the lynx, cat, and tiger phylogeny. **Datasheet S5:** List of putative pseudogenes. **Datasheet S6:** Partitions identified by Saguaro as yielding alternative topologies from a whole genome alignment of Felidae. **Datasheet S7:** List of genes with signatures of positive selection. **Datasheet S8:** List of Gene Ontology terms significantly enriched in windows of high and low diversity in Iberian lynx populations or of high of low difference with respect to Eurasian. (XLSX 340 kb)

Acknowledgements
Iberian lynx samples were contributed by the Iberian lynx Ex Situ Breeding Programme, the Centro de Análisis y Diagnóstico de la Fauna Silvestre, and the LIFE projects LIFE06NAT/E/000209 and LIFE10NAT/ES/570, with authorization of Consejería de Medio Ambiente de la Junta de Andalucía. The Eurasian lynx sample was donated by the Zoológico de Córdoba, Spain. Roche Diagnostics contributed mate-pair 454 sequence data. Centro de Supercomputación de Galicia (CESGA) provided computation and storage services. We thank our NCBI/NIH colleagues Richa Agarwala, Sergey Shiryev, and Aleksandr Morgulis for generating a reference-assisted assembly and its alignment to the domestic cat genome, which encouraged and inspired the assembly approach we finally implemented. We acknowledge support of the publication fee by the CSIC Open Access Publication Support Initiative through its Unit of Information Resources for Research (URICI). We also thank the two anonymous reviewers whose comments and suggestions helped improve this manuscript.

Funding
The financial support for this work was provided by Banco de Santander, CSIC, and Fundación General CSIC through a "Proyectos Cero" grant. FC and BM-C were hired under the EcoGenes project funded by the European FP7 program (264125-FP7-REGPOT-2010-1) and AVl and FCa under the INB Project PT13/0001/0021 (ISCIII - Subdirección General de Evaluación y Fomento de la Investigación/FEDER "Una Manera de hacer Europa"). EBD-CSIC and CRG received support from the Spanish Ministry of Economy and Competitiveness under the "Centro de Excelencia Severo Ochoa 2013-2017" program, grants SEV-2012-0262 and SEV-2012-0208.

Availability of data and materials
The datasets supporting the conclusions of this article are available in the European Nucleotide Archive (ENA; http://www.ebi.ac.uk/ena) as study PRJEB12609. A browser for the Iberian lynx genome assembly is available at http://denovo.cnag.cat/genomes/iberian_lynx.

Authors' contributions
JAG led the project and the manuscript preparation. JAG, JLG, RG, and IG designed and managed the project and together with TA, TG, CN, MMA, TM-B, CL-O, AVa, AR-H, TM, and WJM supervised subprojects/tasks. BAC and fosmid library generation: BG and JLG*. Illumina sequencing production: JB,

MG, IG*. Whole genome assembly: AC, LF, PR, and TA*. Mitochondrial genome assembly: MC-M, AC, JAG, and TA*. Y chromosome assembly and annotation: GL and WJM*. Protein-coding gene annotation: FCa, AC, and TA*. Functional annotation: AVI, GR, EL, and RG*. Splicing isoforms and annotation evaluation: JMR and MT*. Orphan genes: JLV-C, JR-O, and MMA*. Degradome characterization: VQ, JRA, and CL-O*. sncRNA annotation: LC and EL. Gene expression analyses: AVI, FR, JLV-C, MMA, and RG*. lncRNA annotation and expression: PP, IE, and CN*. Eurasian and Iberian lynx divergence: JYC and TM*. Demographic history: FCr*, BM-C*, and JAG. Karyotype characterization: FG, MA-N, and AR-H*. Synteny analyses: FA*, MR-C, BG, and JAG. Phylogenomic analyses: MM-H, SC-G, JLR-A, and TG*; GL and WJM*. Positive selection: JLV-C, JR-O, MM-H, FA, JAG, and MMA* Transposable element annotation and dynamics: FA* and MR-C. Substitution patterns: FA* and FCr. Segmental duplications and structural variation: JQ, BL-G, and TM-B*. Variant discovery and genotyping: FCr*, MR-C, and SD. Population genomics: FCr, FA, BM-C, LS, and JAG*. Lynx genome browser: AC, TA, and PR*. Asterisks indicate task leaders. All authors read and approved the final manuscript.

Competing interests
The authors declare that they have no competing interests.

Ethics approval
We exclusively used samples collected by expert veterinarians during the course of routine check-ups of captive animals, revisions of animals captured for other reasons, or from necropsies. No animals were trapped or sacrificed for the purposes of this study and therefore a formal approval by an Institutional Animal Care and Use Committee was not necessary. Samples were used for this study under permits of the Consejería de Medio Ambiente of the Junta de Andalucía (Andalusian Government).

Author details
[1]Structural Biology and Biocomputing Programme, Spanish National Cancer Research Centre (CNIO), Madrid 28029, Spain. [2]CNAG-CRG, Centre for Genomic Regulation (CRG), Barcelona Institute of Science and Technology (BIST), Baldiri i Reixac 4, 08028 Barcelona, Spain. [3]Department of Integrative Ecology, Doñana Biological Station (EBD), Spanish National Research Council (CSIC), C/ Americo Vespucio, s/n, 41092 Sevilla, Spain. [4]Evolutionary Genomics Group, Research Programme on Biomedical Informatics (GRIB), Hospital del Mar Research Institute (IMIM), Dr. Aiguader 88, 08003 Barcelona, Spain. [5]Bioinformatics and Genomics Programme, Centre for Genomic Regulation (CRG), Dr. Aiguader 88, 08003 Barcelona, Spain. [6]Universitat Pompeu Fabra (UPF), Dr. Aiguader 88, 08003 Barcelona, Spain. [7]Bioinformatics Research Centre, Aarhus University, C.F. Møllers Allé 8, 8000 Aarhus, Denmark. [8]Departamento de Bioquímica y Biología Molecular, Instituto Universitario de Oncología (IUOPA), Universidad de Oviedo, 33006 Oviedo, Spain. [9]Institut de Biologia Evolutiva (UPF-CSIC), Universitat Pompeu Fabra, PRBB, Doctor Aiguader, 88, 08003 Barcelona, Spain. [10]Department of Veterinary Integrative Biosciences, College of Veterinary Medicine, Texas A&M University, College Station, TX 77843, USA. [11]Servei de Cultius Cel.lulars (SCC, SCAC), Universitat Autònoma de Barcelona, Barcelona, Spain. [12]National Bioinformatics Institute (INB), Spanish National Cancer Research Centre (CNIO), Madrid 28029, Spain. [13]Bioinformatics Core Facility, Centre for Genomic Regulation (CRG), Dr. Aiguader 88, 08003 Barcelona, Spain. [14]Department of Environmental Biology, Center for Biological Research (CIB), Spanish National Research Council (CSIC), Ramiro de Maeztu 9, 28040 Madrid, Spain. [15]Institut de Biotecnologia i de Biomedicina, Universitat Autònoma de Barcelona, 08193 Cerdanyola del Vallès, Spain. [16]Computational Genomics Group, Research Programme on Biomedical Informatics (GRIB), Hospital del Mar Research Institute (IMIM), Dr. Aiguader 88, 08003 Barcelona, Spain. [17]Departament de Biologia Cel.lular, Fisiologia i Immunologia, Universitat Autònoma de Barcelona, 08193 Cerdanyola del Vallès, Spain. [18]Institució Catalana de Recerca i Estudis Avançats (ICREA), Pg. Lluís Companys 23, 08010 Barcelona, Spain.

Received: 26 February 2016 Accepted: 25 October 2016
Published online: 14 December 2016

References
1. Pimm SL, Jenkins CN, Abell R, Brooks TM, Gittleman JL, Joppa LN, Raven PH, Roberts CM, Sexton JO. The biodiversity of species and their rates of extinction, distribution, and protection. Science. 2014;344:1246752.
2. WWF. Living Planet Report 2016. Risk and resilience in a new era. Gland: WWF International. 2016:p.145.
3. Thomas CD, Williamson M. Extinction and climate change. Nature. 2012;482:E4–5.
4. Frankham R, Ballou JD, Briscoe DA. Introduction to conservation genetics. 2nd ed. New York: Cambridge University Press; 2010.
5. Shafer ABA, Wolf JBW, Alves PC, Bergstrom L, Bruford MW, Brannstrom I, Colling G, Dalen L, De Meester L, Ekblom R, et al. Genomics and the challenging translation into conservation practice. Trends Ecol Evol. 2015;30:78–87.
6. Li G, Davis BW, Eizirik E, Murphy WJ. Phylogenomic evidence for ancient hybridization in the genomes of living cats (Felidae). Genome Res. 2016;26:1–11.
7. Johnson WE, Eizirik E, Pecon-Slattery J, Murphy WJ, Antunes A, Teeling E, O'Brien SJ. The Late Miocene radiation of modern Felidae: a genetic assessment. Science. 2006;311:73–7.
8. Nowell K, Jackson P, IUCN/SSC Cat Specialist Group. Wild cats: status survey and conservation action plan. Gland: IUCN; 1996.
9. *Lynx pardinus*. The IUCN red list of threatened species. http://www.iucnredlist.org. Accessed 10 Sept 2016.
10. Casas-Marce M, Soriano L, López-Bao JV, Godoy JA. Genetics at the verge of extinction: insights from the Iberian lynx. Mol Ecol. 2013;22:5503–15.
11. Ruiz-López MJ, Gañán N, Godoy JA, Del Olmo A, Garde J, Espeso G, Vargas A, Martínez F, Roldán ERS, Gomendio M. Heterozygosity-fitness correlations and inbreeding depression in two critically endangered mammals. Conserv Biol. 2012;26:1121–9.
12. Palomares F, Godoy JA, López-Bao JV, Rodríguez A, Roques S, Casas-Marce M, Revilla E, Delibes M. Possible extinction vortex for a population of Iberian lynx on the verge of extirpation. Conserv Biol. 2012;26:689–97.
13. Peña L, García P, Jiménez MA, Benito A, Alenza MDP, Sánchez B. Histopathological and immunohistochemical findings in lymphoid tissues of the endangered Iberian lynx (*Lynx pardinus*). Comp Immunol Microbiol Infect Dis. 2006;29:114–26.
14. Jiménez MA, Sánchez B, Alenza MDP, García P, López JV, Rodríguez A, Muñoz A, Martínez F, Vargas A, Peña L. Membranous glomerulonephritis in the Iberian lynx (*Lynx pardinus*). Vet Immunol Immunopathol. 2008;121:34–43.
15. Martínez F, Manteca X, Pastor J. Retrospective study of morbidity and mortality of captive Iberian lynx (*Lynx pardinus*) in the *ex situ* conservation programme (2004-june 2010). J Zoo Wildl Med. 2013;44:845–52.
16. Steiner CC, Putnam AS, Hoeck PEA, Ryder OA. Conservation genomics of threatened animal species. Annu Rev An Biosci. 2013;1:261–81.
17. Kitzman JO, MacKenzie AP, Adey A, Hiatt JB, Patwardhan RP, Sudmant PH, Ng SB, Alkan C, Qiu RL, Eichler EE, Shendure J. Haplotype-resolved genome sequencing of a Gujarati Indian individual. Nat Biotechnol. 2011;29:59–63.
18. Parra G, Bradnam K, Ning Z, Keane T, Korf I. Assessing the gene space in draft genomes. Nucleic Acids Res. 2009;37:289–97.
19. Mailund T, Dutheil JY, Hobolth A, Lunter G, Schierup MH. Estimating divergence time and ancestral effective population size of Bornean and Sumatran orangutan subspecies using a coalescent hidden Markov model. PLoS Genet. 2011;7:e1001319.
20. Mailund T, Halager AE, Westergaard M, Dutheil JY, Munch K, Andersen LN, Lunter G, Prufer K, Scally A, Hobolth A, Schierup MH. A new isolation with migration model along complete genomes infers very different divergence processes among closely related great ape species. PLoS Genet. 2012;8:e1003125.
21. Li H, Durbin R. Inference of human population history from individual whole-genome sequences. Nature. 2011;475:493–6.
22. Gutenkunst RN, Hernandez RD, Williamson SH, Bustamante CD. Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. PLoS Genet. 2009;5:e1000695.
23. Clavero M, Delibes M. Using historical accounts to set conservation baselines: the case of lynx species in Spain. Biodivers Conserv. 2013;22:1691–702.
24. Rodríguez-Varela R, Tagliacozzo A, Ureña I, García N, Crégut-Bonnoure E, Mannino MA, Arsuaga JL, Valdiosera C. Ancient DNA evidence of Iberian lynx palaeoendemism. Quaternary Sci Rev. 2015;112:172–80.
25. Koen EL, Bowman J, Lalor JL, Wilson PJ. Continental-scale assessment of the hybrid zone between bobcat and Canada lynx. Biol Conserv. 2014;178:107–15.
26. Schmidt K, Ratkiewicz M, Konopinski MK. The importance of genetic variability and population differentiation in the Eurasian lynx *Lynx lynx* for conservation, in the context of habitat and climate change. Mamm Rev. 2011;41:112–24.

27. Ellis EC, Goldewijk KK, Siebert S, Lightman D, Ramankutty N. Anthropogenic transformation of the biomes, 1700 to 2000. Global Ecol Biogeogr. 2010;19:589–606.
28. Nie W, Wang J, Su W, Wang D, Tanomtong A, Perelman PL, Graphodatsky AS, Yang F. Chromosomal rearrangements and karyotype evolution in carnivores revealed by chromosome painting. Heredity. 2012;108:17–27.
29. Huerta-Cepas J, Gabaldón T. Assigning duplication events to relative temporal scales in genome-wide studies. Bioinformatics. 2011;27:38–45.
30. Lombardi G, Burzyn D, Mundinano J, Berguer P, Bekinschtein P, Costa H, Castillo LF, Goldman A, Meiss R, Piazzon I, Nepomnaschy I. Cathepsin-L influences the expression of extracellular matrix in lymphoid organs and plays a role in the regulation of thymic output and of peripheral T cell number. J Immunol. 2005;174:7022–32.
31. Schurigt U, Eilenstein R, Gajda M, Leipner C, Sevenich L, Reinheckel T, Peters C, Wiederanders B, Brauer R. Decreased arthritis severity in cathepsin L-deficient mice is attributed to an impaired T helper cell compartment. Inflammation Res. 2012;61:1021–9.
32. Yang ZH. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol. 2007;24:1586–91.
33. Heffner HE, Heffner RS. The sound-localization ability of cats. J Neurophysiol. 2005;94:3653–3.
34. Hunt DM, Buch P, Michaelides M. Guanylate cyclases and associated activator proteins in retinal disease. Mol Cell Biochem. 2010;334:157–68.
35. Boutin TS, Le Rouzic A, Capy P. How does selfing affect the dynamics of selfish transposable elements? Mob DNA. 2012;3:5.
36. Lynch M, Conery JS. The origins of genome complexity. Science. 2003;302:1401–4.
37. Morgan MT. Transposable element number in mixed mating populations. Genet Res. 2001;77:261–75.
38. Gherman A, Chen PE, Teslovich TM, Stankiewicz P, Withers M, Kashuk CS, Chakravarti A, Lupski JR, Cutler DJ, Katsanis N. Population bottlenecks as a potential major shaping force of human genome architecture. PLoS Genet. 2007;3:e119.
39. Lockton S, Ross-Ibarra J, Gaut BS. Demography and weak selection drive patterns of transposable element diversity in natural populations of Arabidopsis lyrata. Proc Natl Acad Sci U S A. 2008;105:13965–70.
40. Zhang Y, Romanish MT, Mager DL. Distributions of transposable elements reveal hazardous zones in mammalian introns. PLoS Comp Biol. 2011;7:e1002046.
41. Glemin S, Arndt PF, Messer PW, Petrov D, Galtier N, Duret L. Quantification of GC-biased gene conversion in the human genome. Genome Res. 2015;25:1215–28.
42. Katzman S, Kern AD, Pollard KS, Salama SR, Haussler D. GC-biased evolution near human accelerated regions. PLoS Genet. 2010;6:e1000960.
43. Cho YS, Hu L, Hou H, Lee H, Xu J, Kwon S, Oh S, Kim H-M, Jho S, Kim S, et al. The tiger genome and comparative analysis with lion and snow leopard genomes. Nat Commun. 2013;4:2433.
44. Li S, Li B, Cheng C, Xiong Z, Liu Q, Lai J, Carey H, Zhang Q, Zheng H, Wei S, et al. Genomic signatures of near-extinction and rebirth of the crested ibis and other endangered bird species. Genome Biol. 2014;15:557.
45. Zhao S, Zheng P, Dong S, Zhan X, Wu Q, Guo X, Hu Y, He W, Zhang S, Fan W, et al. Whole-genome sequencing of giant pandas provides insights into demographic history and local adaptation. Nat Genet. 2013;45:67–71.
46. Xue Y, Prado-Martínez J, Sudmant PH, Narasimhan V, Ayub Q, Szpak M, Frandsen P, Chen Y, Yngvadottir B, Cooper DN, et al. Mountain gorilla genomes reveal the impact of long-term population decline and inbreeding. Science. 2015;348:242–5.
47. Der Sarkissian C, Ermini L, Schubert M, Yang Melinda A, Librado P, Fumagalli M, Jónsson H, Bar-Gal Gila K, Albrechtsen A, Vieira Filipe G, et al. Evolutionary genomics and conservation of the endangered Przewalski's horse. Curr Biol. 2015;25:2577–83.
48. Alhaddad H, Khan R, Grahn RA, Gandolfi B, Mullikin JC, Cole SA, Gruffydd-Jones TJ, Haggstrom J, Lohi H, Longeri M, Lyons LA. Extent of linkage disequilibrium in the domestic cat, *Felis silvestris catus*, and its breeds. Plos One. 2013;8:e53537.
49. Hill WG, Robertson A. The effect of linkage on limits to artificial selection. Genet Res. 1966;8:269–94.
50. Leffler EM, Bullaughey K, Matute DR, Meyer WK, Segurel L, Venkat A, Andolfatto P, Przeworski M. Revisiting an old riddle: what determines genetic diversity levels within species? PLoS Biol. 2012;10:e1001388.
51. Bustamante CD, Fledel-Alon A, Williamson S, Nielsen R, Hubisz MT, Glanowski S, Tanenbaum DM, White TJ, Sninsky JJ, Hernandez RD, et al. Natural selection on protein-coding genes in the human genome. Nature. 2005;437:1153–7.
52. Loire E, Chiari Y, Bernard A, Cahais V, Romiguier J, Nabholz B, Lourenco JM, Galtier N. Population genomics of the endangered giant Galapagos tortoise. Genome Biol. 2013;14:R136.
53. Pool JE, Nielsen R. Population size changes reshape genomic patterns of diversity. Evolution. 2007;61:3001–6.
54. Otto SP, Barton NH. Selection for recombination in small populations. Evolution. 2001;55:1921–31.
55. Alonso S, López S, Izagirre N, de la Rua C. Overdominance in the human genome and olfactory receptor activity. Mol Biol Evol. 2008;25:997–1001.
56. Hamilton JA, Miller JM. Adaptive introgression as a resource for management and genetic conservation in a changing climate. Conserv Biol. 2016;30:33–41.
57. Thomas MA, Roemer GW, Donlan CJ, Dickson BG, Matocq M, Malaney J. Gene tweaking for conservation. Nature. 2013;501:485–6.
58. Kovach RP, Luikart G, Lowe WH, Boyer MC, Muhlfeld CC. Risk and efficacy of human-enabled interspecific hybridization for climate-change adaptation: response to Hamilton and Miller (2016). Conserv Biol. 2016;30:428–30.
59. Reed DH. Albatrosses, eagles and newts, Oh My!: exceptions to the prevailing paradigm concerning genetic diversity and population viability? Anim Conserv. 2010;13:448–57.
60. Palomares F, Rodríguez A, Revilla E, López-Bao JV, Calzada J. Assessment of the conservation efforts to prevent extinction of the *Iberian lynx*. Conserv Biol. 2011;25:4–8.
61. Griebel T, Zacher B, Ribeca P, Raineri E, Lacroix V, Guigo R, Sammeth M. Modelling and simulating generic RNA-Seq experiments with the flux simulator. Nucleic Acids Res. 2012;40:10073–83.
62. Ramskold D, Wang ET, Burge CB, Sandberg R. An abundance of ubiquitously expressed genes revealed by tissue transcriptome sequence data. PLoS Comp Biol. 2009;5:e1000598.
63. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010;26:139–40.
64. Brawand D, Soumillon M, Necsulea A, Julien P, Csardi G, Harrigan P, Weier M, Liechti A, Aximu-Petri A, Kircher M, et al. The evolution of gene expression levels in mammalian organs. Nature. 2011;478:343–8.
65. Tarazona S, Garcia-Alcalde F, Dopazo J, Ferrer A, Conesa A. Differential expression in RNA-seq: a matter of depth. Genome Res. 2011;21:2213–23.
66. RepeatMasker Open-4.0. http://www.repeatmasker.org.
67. Morgulis A, Gertz EM, Schäffer AA, Agarwala R. A fast and symmetric DUST implementation to mask low-complexity DNA sequences. J Comput Biol. 2006;13:1028–40.
68. Haas BJ, Salzberg SL, Zhu W, Pertea M, Allen JE, Orvis J, White O, Buell CR, Wortman JR. Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. Genome Biol. 2008;9:R7.
69. Nawrocki EP, Kolbe DL, Eddy SR. Infernal 1.0: inference of RNA alignments. Bioinformatics. 2009;25:1335–7.
70. Griffiths-Jones S, Moxon S, Marshall M, Khanna A, Eddy SR, Bateman A. Rfam: annotating non-coding RNAs in complete genomes. Nucleic Acids Res. 2005;33:D121–4.
71. Esteve-Codina A, Kofler R, Palmieri N, Bussotti G, Notredame C, Pérez-Enciso M. Exploring the gonad transcriptome of two extreme male pigs with RNA-seq. BMC Genomics. 2011;12:552.
72. Derrien T, Johnson R, Bussotti G, Tanzer A, Djebali S, Tilgner H, Guernec G, Martin D, Merkel A, Knowles DG, et al. The GENCODE v7 catalog of human long noncoding RNAs: Analysis of their gene structure, evolution, and expression. Genome Res. 2012;22:1775–89.
73. Zdobnov EM, Apweiler R. InterProScan–an integration platform for the signature-recognition methods in InterPro. Bioinformatics. 2001;17:847–8.
74. Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M. KEGG for integration and interpretation of large-scale molecular data sets. Nucleic Acids Res. 2012;40:D109–14.
75. Matthews L, Gopinath G, Gillespie M, Caudy M, Croft D, de Bono B, Garapati P, Hemish J, Hermjakob H, Jassal B, et al. Reactome knowledgebase of human biological pathways and processes. Nucleic Acids Res. 2009;37:D619–22.
76. Gotz S, Garcia-Gómez JM, Terol J, Williams TD, Nagaraj SH, Nueda MJ, Robles M, Talon M, Dopazo J, Conesa A. High-throughput functional annotation and data mining with the Blast2GO suite. Nucleic Acids Res. 2008;36:3420–35.

77. Petersen TN, Brunak S, von Heijne G, Nielsen H. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat Methods. 2011;8:785–6.
78. Ordoñez GR, Puente XS, Quesada V, López-Otín C. Proteolytic systems: constructing degradomes. Met Mol Biol. 2009;539:33–47.
79. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990;215:403–10.
80. Toll-Riera M, Bosch N, Bellora N, Castelo R, Armengol L, Estivill X, Albà MM. Origin of primate orphan genes: A comparative genomics approach. Mol Biol Evol. 2009;26:603–12.
81. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol. 2010;28:511–5.
82. Seabright M. Use of proteolytic-enzymes for mapping of structural rearrangements in chromosomes of man. Chromosoma. 1972;36:204.
83. Wursterhill DH, Centerwall WR. The interrelationships of chromosome-banding patterns in canids, mustelids, hyena, and felids. Cytogenet Cell Genet. 1982;34:178–92.
84. O'Brien JT, Menninger JC, Nash WG. Atlas of mammalian chromosomes. Hoboken: Wiley; 2006.
85. Kiełbasa SM, Wan R, Sato K, Horton P, Frith MC. Adaptive seeds tame genomic sequence comparison. Genome Res. 2011;21:487–93.
86. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010;26:841–2.
87. Huerta-Cepas J, Capella-Gutiérrez S, Pryszcz LP, Denisov I, Kormes D, Marcet-Houben M, Gabaldón T. PhylomeDB v3.0: an expanding repository of genome-wide collections of trees, alignments and phylogeny-based orthology and paralogy predictions. Nucleic Acids Res. 2011;39:D556–60.
88. Gil M, Zanetti MS, Zoller S, Anisimova M. CodonPhyML: fast maximum likelihood phylogeny estimation under codon substitution models. Mol Biol Evol. 2013;30:1270–80.
89. Huerta-Cepas J, Dopazo J, Gabaldón T. ETE: a python environment for tree exploration. BMC Bioinf. 2010;11:24–4.
90. Huerta-Cepas J, Dopazo H, Dopazo J, Gabaldón T. The human phylome. Genome Biol. 2007;8:R109–9.
91. Huerta-Cepas J, Capella-Gutiérrez S, Pryszcz LP, Marcet-Houben M, Gabaldón T. PhylomeDB v4: zooming into the plurality of evolutionary histories of a genome. Nucleic Acids Res. 2014;42:D897–902.
92. Al-Shahrour F, Minguez P, Tárraga J, Medina I, Alloza E, Montaner D, Dopazo J. FatiGO +: a functional profiling tool for genomic data. Integration of functional annotation, regulatory motifs and interaction data with microarray experiments. Nucleic Acids Res. 2007;35:W91–W96.
93. Slater GS, Birney E. Automated generation of heuristics for biological sequence comparison. BMC Bioinf. 2005;6:e31.
94. Zamani N, Russell P, Lantz H, Hoeppner MP, Meadows JRS, Vijay N, Mauceli E, di Palma F, Lindblad-Toh K, Jern P, Grabherr MG. Unsupervised genome-wide recognition of local relationship patterns. BMC Genomics. 2013;14:347.
95. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014;30:1312–3.
96. Loytynoja A, Goldman N. An algorithm for progressive multiple alignment of sequences with insertions. Proc Natl Acad Sci U S A. 2005;102:10557–62.
97. Villanueva-Cañas JL, Laurie S, Albà MM. Improving genome-wide scans of positive selection by using protein isoforms of similar length. Genome Biol Evol. 2013;5:457–67.
98. Pérez-Llamas C, López-Bigas N. Gitools: Analysis and visualisation of genomic data using interactive heat-maps. PLoS One. 2011;6:e19541.
99. Flicek P, Ahmed I, Amode MR, Barrell D, Beal K, Brent S, Carvalho-Silva D, Clapham P, Coates G, Fairley S, et al. Ensembl 2013. Nucleic Acids Res. 2013;41:D48–55.
100. Gascuel O. BIONJ: An improved version of the NJ algorithm based on a simple model of sequence data. Mol Biol Evol. 1997;14:685–95.
101. Rubio-Camarillo M, Gómez-López G, Fernández JM, Valencia A, Pisano DG. RUbioSeq: a suite of parallelized pipelines to automate exome variation and bisulfite-seq analyses. Bioinformatics. 2013;29:1687–9.
102. Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, Land SJ, Lu XY, Ruden DM. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w(1118); iso-2; iso-3. Fly. 2012;6:80–92.
103. Rodríguez JM, Maietta P, Ezkurdia I, Pietrelli A, Wesselink J-J, López G, Valencia A, Tress ML. APPRIS: annotation of principal and alternative splice isoforms. Nucleic Acids Res. 2013;41:D110–7.
104. Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics. 2009;25:1754–60.
105. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Proc GPD. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009;25:2078–9.
106. Hach F, Sarrafi I, Hormozdiari F, Alkan C, Eichler EE, Sahinalp SC. mrsFAST-Ultra: a compact, SNP-aware mapper for high performance sequencing applications. Nucleic Acids Res. 2014;42:W494–500.
107. Alkan C, Kidd JM, Marques-Bonet T, Aksay G, Antonacci F, Hormozdiari F, Kitzman JO, Baker C, Malig M, Mutlu O, et al. Personalized copy number and segmental duplication maps using next-generation sequencing. Nat Genet. 2009;41:1061–7.
108. Iqbal Z, Caccamo M, Turner I, Flicek P, McVean G. De novo assembly and genotyping of variants using colored de Bruijn graphs. Nat Genet. 2012;44:226–32.
109. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, et al. The variant call format and VCFtools. Bioinformatics. 2011;27:2156–8.
110. Davinson AC, Hinkley DV. Bootstrap methods and their application. Cambridge: Cambridge University Press; 1997.
111. Canty A, Ripley B. boot: Bootstrap R (S-Plus) Functions. R package version 1. 3-11. 2014.
112. Al-Shahrour F, Díaz-Uriarte R, Dopazo J. FatiGO: a web tool for finding significant associations of Gene Ontology terms with groups of genes. Bioinformatics. 2004;20:578–80.
113. Medina I, Carbonell J, Pulido L, Madeira SC, Goetz S, Conesa A, Tarraga J, Pascual-Montano A, Nogales-Cadenas R, Santoyo J, et al. Babelomics: an integrative platform for the analysis of transcriptomics, proteomics and genomic data with advanced functional profiling. Nucleic Acids Res. 2010;38:W210–3.
114. Prufer K, Racimo F, Patterson N, Jay F, Sankararaman S, Sawyer S, Heinze A, Renaud G, Sudmant PH, de Filippo C, et al. The complete genome sequence of a Neanderthal from the Altai Mountains. Nature. 2014;505:43–9.
115. Dobrynin P, Liu S, Tamazian G, Xiong Z, Yurchenko A, Krasheninnikova K, Kliver S, Schmidt-Kuntzel A, Koepfli K-P, Johnson W, et al. Genomic legacy of the African cheetah, Acinonyx jubatus. Genome Biol. 2015;16:277.
116. Zhou XM, Sun FM, Xu SX, Fan GY, Zhu KL, Liu X, Chen Y, Shi CC, Yang YX, Huang ZY, et al. Baiji genomes reveal low genetic variability and new insights into secondary aquatic adaptations. Nat Commun. 2013;4:2708.
117. Carbone L, Alan Harris R, Gnerre S, Veeramah KR, Lorente-Galdos B, Huddleston J, Meyer TJ, Herrero J, Roos C, Aken B, et al. Gibbon genome and the fast karyotype evolution of small apes. Nature. 2014;513:195–201.
118. Yim HS, Cho YS, Guang XM, Kang SG, Jeong JY, Cha SS, Oh HM, Lee JH, Yang EC, Kwon KK, et al. Minke whale genome and aquatic adaptation in cetaceans. Nat Genet. 2014;46:88–92.
119. Wu HG, Guang XM, Al-Fageeh MB, Cao JW, Pan SK, Zhou HM, Zhang L, Abutarboush MH, Xing YP, Xie ZY, et al. Camelid genomes reveal evolution and adaptation to desert environments. Nat Commun. 2014;5:5188.
120. Palkopoulou E, Mallick S, Skoglund P, Enk J, Rohland N, Li H, Omrak A, Vartanyan S, Poinar H, Götherström A, et al. Complete genomes reveal signatures of demographic and genetic declines in the woolly mammoth. Curr Biol. 2015;25:1395–400.
121. Fang XD, Nevo E, Han LJ, Levanon EY, Zhao J, Avivi A, Larkin D, Jiang XT, Feranchuk S, Zhu YB, et al. Genome-wide adaptive complexes to underground stresses in blind mole rats Spalax. Nat Commun. 2014;5:3966.