

Studie zur genetischen Diversität und Populationsstruktur des Leonbergers

Anna Letko¹, Katie M. Minor², Vidhya Jagannathan¹, Franz R. Seefried³, James R. Mickelson², Pieter Oliehoek⁴, and Cord Drögemüller¹

¹ Institute of Genetics, Vetsuisse Faculty, University of Bern, 3012 Bern, Switzerland

² Department of Veterinary and Biomedical Sciences, College of Veterinary Medicine, University of Minnesota, St. Paul, Minnesota, 55108, USA

³ Qualitas AG, 6300 Zug, Switzerland

⁴ Dogs Global, Wageningen, the Netherlands

Zusammenfassung

Hintergrund Der Leonberger gehört zu den grossen Hunderassen, die in den 1850er Jahren in Deutschland entstanden. Seine Beliebtheit hat nach dem Zweiten Weltkrieg zu seiner aktuellen Population von weltweit rund 30.000 Hunden geführt.

Die Rasse zeigt neurodegenerative Erkrankungen und Krebserkrankungen, was mit grosser Wahrscheinlichkeit auf die begrenzte genetische Vielfalt zurückzuführen ist.

Es gibt jedoch keinen wissenschaftlichen Bericht über Parameter der Genetik und die sogenannte genomische Architektur der heutigen Population.

Resultate: Wir haben hierfür umfangreiche Stammbaumaufzeichnungen von 142.072 Leonbergern, sowie umfassende DNA-Analysedaten, genauer sogenannte SNP-Array-Genotypdaten von 1.203 Hunden und Ganzgenomsequenzen (WGS Whole genome sequencing) von 39 Hunden, ausgewertet. Die Stammbaumanalysen identifizierten 22 Gründertiere und zeigten einen offensichtlichen Vererber-, Rüden-, oder Vätereffekt (*Der beliebte Vater-Effekt (oder das beliebte Deck-/Vater-Syndrom) tritt auf, wenn ein Tier mit wünschenswerten Eigenschaften wiederholt gezüchtet wird. In der Hundezucht ist ein Rüde, der angesehene Wettbewerbe gewinnt, sehr gefragt, da die Züchter glauben, dass der Vater die Gene besitzt, die für die Produktion von Champions erforderlich sind. Der beliebte Rüde deckt viele Hündinnen und produziert viele Nachkommen. Dies kann dazu führen, dass sich unentdeckte, unerwünschte genetische Merkmale in der Zucht schnell innerhalb des Genpools ausbreiten. Der Vätereffekt kann auch die genetische Vielfalt durch den Ausschluss anderer Rüden verringern*).

Der durchschnittliche stammbaumbasierte Inzuchtkoeffizient von 0,29 sowie die durchschnittliche Verwandtschaft (Meankinship) von 0,31 zeigten einen dramatischen Verlust der genetischen Vielfalt. Die beobachtete durchschnittliche Lebenserwartung sank im Laufe der Zeit von 9,4 Jahren im Jahr 1989 auf 7,7 Jahre im Jahr 2004.

Eine globale Umfrage zur Gesundheit bestätigte eine hohe Prävalenz (Krankheitshäufigkeit) von Krebs sowie neurologischen Erkrankungen.

Die SNP-basierte Analyse der Segmente der Homozygotie (ROH) identifizierte 125'653 ROHs mit einer durchschnittlichen Länge von 5,88 Mb und bestätigte einen durchschnittlichen Inzuchtkoeffizienten von 0,28 (*d.h. umfangreiche Abschnitte des Leonberger Genoms sind auf beiden DNA-Strängen abstammungsidentisch*).

Die genomweite Filterung der WGS-Daten ergab 28 nicht-kodierende Genvarianten, die in allen Leonbergern vorkommen, und eine Liste von 22 potentiell pathogenen (krankheitsauslösenden) Genvarianten für neurologische Erkrankungen; die Hälfte dieser Varianten traten nur in Leonberger Hunden auf, während die andere Hälfte der Varianten mit seltener Häufigkeit in weiteren verschiedenen Hunderassen zu finden waren. Außerdem war eine der beiden gefundenen sogenannten Haplogruppen der maternal vererbten mitochondrialen DNA nur bei einem von 39 Hunden vorhanden.

Schlussfolgerung: Die zunehmende Größe der Leonberger-Population ging nach dem Flaschenhalseffekt (bottleneck-event in den 1940er Jahren (*im 2. Weltkrieg brach die Hundezucht und Hundepopulation zusammen*)) mit einem beträchtlichen Verlust an genetischer Vielfalt einher. Die intensive Verwendung beliebter Zuchtrüden (*Vätereffekt*) führte in einem hohen Masse zur Steigerung der Inzucht der Rasse.

Dies könnte auch die hohe Prävalenz bestimmter Erkrankungen erklären; die genomischen Daten liefern jedoch keine Hinweise auf fixe kodierende Genvarianten, also Mutationen die bei nahezu allen Leonbergern gleich vorkommen, die diese rassespezifischen Prädispositionen erklären. Die Liste der möglichen ursächlichen Varianten für weitere bislang ungeklärte Formen der häufig auftretenden Polyneuropathie beim Leonberger muss weiter ausgewertet werden. **Die Erhaltung der aktuellen genetischen Vielfalt ist möglich, indem (1.) grundsätzlich mehr Hunde aktiv in die Zucht einbezogen werden, während (2.) die Anzahl der Würfe pro Vater/Mutter klar eingeschränkt bzw. begrenzt wird.**

Zusätzlich würde eine Auskreuzung (Kreuzung verschiedener Rassen, die keine gemeinsame Vorfahren besitzen) helfen, langfristig die genetische Vielfalt zu erhöhen und somit zu optimieren und zur Nachhaltigkeit und Gesundheit der Population beizutragen.

Einleitung

Der Leonberger gehört zu den Riesen-Hunde-Rassen, die um die 1850er Jahre in Deutschland als Wach-, Begleit- und Familienhunde entstanden [1]. Die Rasse wurde nach dem Zweiten Weltkrieg immer beliebter, was zu einer geschätzten aktuellen Population von ~30.000 Hunden führte [2]. Der Leonberger scheint eine höhere Veranlagung für neurodegenerative Erkrankungen und Krebsarten wie Hämangiosarkome und Osteosarkome zu haben als andere Hunderassen [3]. Neuere Untersuchungen haben gezeigt, dass die Genvarianten ARHGEF10 [4] **LPN1** und GJA9 [5] **LPN2** etwa ein Drittel der mit Polyneuropathie diagnostizierten Leonberger-Fälle erklären (OMIA 001917-9615; OMIA 002119-9615).

Eine rezessiv vererbte Genvariante NAPEPLD [6] verursacht eine juvenile Leukoencephalomyelopathie **LEMP** (OMIA 001788-9615).

In der Vergangenheit stützte sich die Schätzung der Inzucht auf detaillierte Stammbäume, wobei ein Inzuchtkoeffizient aus stammbaumbasierten Beziehungen zwischen Vorfahren (F_PED) geschätzt wurde.

Polymorphe Mikrosatellitenmarker-Genotypdaten wurden bereits früher zur Bewertung der genetischen Diversität bei Hunden verwendet (z. B. [7, 8]), wobei genomweite Single Nucleotide Polymorphism (SNP)-Genotypdaten in jüngerer Zeit bei einer Reihe von Rassen verwendet wurden [9-13]. SNP-Daten ermöglichen auch die Charakterisierung von Runs of Homozygosity (ROH), die eine Quantifizierung des Ausmaßes der Inzucht bei diploiden Individuen ermöglichen (*Diploide Zellen weisen einen doppelten Chromosomensatz (46) auf. Die meisten Säugetier- Zellen sind diploid*), insbesondere im Fall von unvollständigen, unzuverlässigen oder fehlenden

Stammbauminformationen [14]. **Darüber hinaus kann die Länge der beobachteten ROH-Segmente zur Unterscheidung zwischen heutiger (gegenwärtiger) und alter Inzucht verwendet werden [15, 16] und erlaubt Rückschlüsse auf die Populationsgeschichte von Populationen bzw. Rassen [17].**

Seit mehr als einem Jahrzehnt generieren Viehzüchter massive Genotypdaten, während sie genomische Selektionsschemata einführen, während bei Hunden solche Daten hauptsächlich für Genkartierungszwecke generiert werden [18, 19]. Die erste berichtete Verwendung von SNP-Daten zur Untersuchung der Verknüpfung (Assoziation) zwischen dem Maß der reproduktiven Fitness und dem Grad der Inzucht bei Hunden wurde bei Golden Retrievern durchgeführt und zeigte, dass ein statistisch signifikantes negatives Verhältnis (Korrelation) zwischen Fruchtbarkeit (Fekundität) und F_ROH (Homozygotie) besteht [20] (*d.h. je weniger Variabilität im Genom vorhanden ist, desto schlechter ist die Reproduktionsfähigkeit des Hundes*). Im Allgemeinen basieren andere Studien auf SNP-Genotypisierungsdaten mit variabler Markerdichte von meist weniger als 100 ausgewählten Individuen pro Rasse; so wurden z. B. nur 34 Individuen des Deutschen Kurzhaar-Pointers analysiert, um über einen genomischen Inzuchtkoeffizienten basierend auf F_ROH von 0,17 zu berichten, während ein mittlerer Inzuchtkoeffizient von 0,023 basierend auf genealogischen Informationen gefunden wurde, was zeigt, dass letztere unvollständig waren [12] (*in diesem Fall ist der stammbuchbasierte Inzuchtkoeffizient kleiner als der genomische*).

Die derzeit am besten untersuchte Hunderasse, bei der SNP- und Ganzgenomsequenzdaten (WGS) zur Bewertung der genetischen Diversität verwendet werden, ist der Norwegische Lundehund, von dem bekannt ist, dass er ein Risiko für ein rassespezifisches, multifaktorielles, lebensbedrohliches

Syndrom aufweist [21]. Die derzeitige Lunde Hundpopulation ist hoch ingezüchtet und eine optimale Anpaarungsplanung allein ermöglicht aufgrund der extrem hohen Verwandtschaft der gesamten Population keine Verbesserung der aktuellen Situation [22]. Analysen der SNP-Genotypdaten ergaben eine wesentlich geringere genetische Diversität für den Lunde Hund, weshalb zur Rettung der gefährdeten Population eine Auskreuzung mit nahe verwandten Rassen empfohlen wurde [13, 22]. Ein erfolgreiches Beispiel für das Ersetzen eines pathogenen Allels durch Auskreuzen ist die Rasse Dalmatiner, die an Hyperurikosurie leidet und eine extrem hohe Frequenz des deleteren (*schädlichen*) rezessiven Allels des SLC2A9-Urat-Transporter-Gens aufweist. Hier wurde ein Individuum aus der Pointer-Rasse, das homozygot für das Wildtyp-Allel war, verwendet, um nicht betroffene heterozygote Hunde zu erzeugen, die anschließend rückgekreuzt wurden, um gesunde Individuen zu erzeugen, die von reinrassigen Dalmatinern kaum zu unterscheiden sind [23]. WGS (Ganzgenomsequenzierung) hat das Erstellen eines großen genetischen Variantenkatalogs ermöglicht, der einen Großteil der bei modernen Hunden vorhandenen Variationen erfasst [19, 24]. Ein umfassender Satz von Varianten zusammen mit ihren Allel- und Genotyphäufigkeiten innerhalb und zwischen den Rassen hilft, funktionell relevante von neutralen Varianten zu unterscheiden und potenzielle rassespezifische Variationsbereiche zu identifizieren. Diese Daten haben sich als sehr nützlich für die Analyse der phänotypischen Variationen (*Merkmalsausprägung*) erwiesen [25] und haben das Potenzial Loci zu identifizieren, die sowohl zur einfachen als auch zur komplexen Krankheitsanfälligkeit bei Hunden beitragen [26, 27]. Die daraus resultierende Identifizierung von ursächlichen Varianten (*für Gesundheitsprobleme*) trägt zur nachhaltigen Erhaltung von Rassen durch die Reduktion von vererbten Gesundheitsproblemen bei.

Das mitochondriale Genom (mtDNA) (*Die mitochondriale DNA oder kurz mtDNA befindet sich in den Mitochondrien. Das sind Zellorganellen, die für den Energiehaushalt der Zelle verantwortlich sind. Besonders wichtig ist, dass Mitochondrien die einzigen Organellen (außer dem Zellkern) sind, die eine eigene DNA besitzen. Es befindet sich außerhalb der Chromosomen einer Zelle. Die Mitochondrien und damit alles was drin ist, werden nur über die Mutter an die Nachkommen weitergegeben. Alle Nachkommen einer Mutter haben also die gleiche mtDNA. Nur sie wird vererbt. Die Mutter hat ihrerseits dieselbe mtDNA wie ihre Großmutter und Urgroßmutter usw. Die mitochondriale DNA eignet sich also besonders gut, die mütterliche Linie zurückzuverfolgen. Über eine mtDNA Untersuchung kann z. B. festgestellt werden, wie nah zwei Völkergruppen miteinander verwandt sind*), das während der WGS sequenziert wird, kann ebenfalls als zusätzlicher Indikator zur Beurteilung der Diversität verwendet werden, um ein umfassenderes Bild der Herkunft und Geschichte von Hundepopulationen zu erhalten [28-30].

Es gibt bisher keinen Bericht über die genetische Charakterisierung der Rasse Leonberger anhand genomischer Daten. In diesem Bericht haben wir umfangreiche Stammbaumdaten von 142.072 Hunden sowie SNP-Array-Genotypen von 1.203 Hunden und WGS-Daten von 39 Hunden verwendet, um die genetische Vielfalt innerhalb der aktuellen weltweiten Leonberger-Rassepopulation zu bewerten.

Material und Methoden

Tiere

Für diese Studie wurden weltweit Blutproben von 1.203 Leonbergern verwendet (Zusatzdatei 1). Die Genomische DNA wurde entweder mit dem Genra PureGene blood kit (Qiagen) oder dem Maxwell RSC whole blood DNA kit (Promega) aus den Blutproben isoliert. Insgesamt 39 Leonberger Hunde wurden entweder im Rahmen früherer Studien [5, 6, 31] oder als ungeklärte Fälle mit neurologischen Störungen ganzgenomisch sequenziert (Zusatzdatei 1). Stammbaumeinträge waren für 142.072 Hunde über die Worldwide Independent Leonberger Database [2] verfügbar. Gesundheitsdaten waren für 2.726 Hunde verfügbar, die von ihren Besitzern über einen Online-Fragebogen eingereicht worden waren.

Stammbaum-Analysen

Wir haben umfangreiche Stammbaumaufzeichnungen von 142.072 Tieren, die bis 2016 geboren wurden, ausgewertet. Das älteste aufgezeichnete Geburtsdatum datiert aus dem Jahr 1880. Stammbaumanalysen wurden mit der Open-Source-Software EVA v3.0 durchgeführt, [32] sowie mit

der hauseigenen Qualitas-Stammbaumsoftware, die intensiv für routinemäßige genetische Auswertungsanalysen verwendet wird. Der Verlust an genetischer Diversität sowie die Möglichkeit, die Diversität zu erhöhen, wurde anhand der mittleren Verwandtschaft (*MK Meankinship*) mit der tabellarischen Methode bewertet, wie zuvor von Oliehoek et al. beschrieben [33]. MK der aktuellen Population wurde für 31.832 ausgewählte Hunde berechnet, von denen angenommen wird, dass sie für die Zucht zur Verfügung stehen und zum Zeitpunkt der Berechnung am Leben sind. Bei vor 2009 geborenen Rüden und vor 2011 geborenen Hündinnen wurde davon ausgegangen, dass sie nicht mehr in der Zucht sind, sie wurden zusammen mit den verstorbenen Hunden ausgeschlossen. Die tatsächliche Populationsgröße könnte kleiner sein, da der aktuelle Status für viele Tiere unbekannt ist.

Globale Gesundheitsumfrage

Ein Online-Fragebogen wurde an die Mitglieder verschiedener europäischer und amerikanischer Leonberger-Zuchtvereine verschickt und darüber hinaus auch periodisch über Facebook beworben. Ein Link zu einer internetbasierten Version des Fragebogens in sieben Sprachen wurde bereitgestellt, um eine weltweite Reichweite von Besitzern und Züchtern sicher zu stellen [34]. Der Fragebogen besteht aus drei Hauptteilen, in denen allgemeine Informationen über den Hund, Anzeichen für spezifische Krankheiten wie neurologische Störungen und Krebs sowie die gesamte medizinische Vorgeschichte der einzelnen Hunde einschließlich des Todesdatums abgefragt werden. Insgesamt wurden die Antworten von 2.726 Hunden, die zwischen 2013 und 2019 gesammelt wurden, analysiert.

SNP-Array-Genotypisierung und ROH-Analysen

Von den 1.203 Leonbergern mit SNP-Array-Genotypdaten wurden 308 auf dem 460k Axiom Canine Genotyping Array Set A (Thermo Fisher Scientific) und 895 auf dem 170k Illumina CanineHD BeadChip (Illumina) genotypisiert. Die PLINK v1.9 Software [35] wurde verwendet, um die Genotypdaten zusammenzuführen und das Trimmen zur Qualitätskontrolle sowie die Segmente der Homozygotie (ROH) Analysen durchzuführen. Das Zusammenführen der verschiedenen Datensätze beschränkte sich auf überlappende Marker zwischen beiden Arrays und auf Marker wo nicht mehr als 10% der Genotypdaten fehlten (missingness <0,1). Der endgültige Datensatz bestand aus 137'476 SNP-Markern. Zusätzlich wurden nur bi-allelische SNPs auf den 38 Hundeaautosomen für die ROH-Analysen beibehalten und SNPs, die vom Hardy-Weinberg-Gleichgewicht abwichen ($p < 0,0001$), wurden ausgeschlossen. Schließlich standen 132'711 Marker als endgültiger Datensatz mit einem SNP pro 16,52kb zur Verfügung. Eine Mindestanzahl von 70 SNPs und mindestens 1 SNP pro 20 kb, das zur Identifizierung eines ROH erforderlich ist, wurden mit der zuvor beschriebenen Methode [36] berechnet, um <5% zufällig generierter ROH zu erzeugen. Andere ROH-Parameter wurden auf dem Standardwert gehalten, wie er in PLINK [35] definiert ist. Ein auf ROH basierender Inzuchtkoeffizient (F_{ROH}) wurde berechnet als die Gesamtlänge aller ROHs in einem Hund geteilt durch die Gesamtlänge des autosomalen Genoms, das von den verwendeten SNP-Markern abgedeckt wird. Die multidimensionale Skalierung (MDS) der paarweisen genetischen Abstände in PLINK [35] wurde zur Ermittlung der Populationsstruktur verwendet. Alle Genompositionen beziehen sich auf die CanFam3.1-Referenzsequenz. Die Abbildungen wurden in der R-Umgebung v3.6.0 [37] mit den Paketen CMplot v3.5.1 [38], plotrix v 3.7-6 [39] und qqman v0.1.4 [40] gezeichnet.

Whole-genome sequencing und variant calling

Die WGS-Daten von 39 Leonbergern wurden nach der Präparation einer PCR-freien Fragmentbibliothek gewonnen, um eine durchschnittliche 18,1× Abdeckung (von 8,3× bis 33,9×; Zusatzdatei 1) zu erzeugen, wie zuvor beschrieben [24]. Fastq-Dateien wurden auf die Hunde-Referenzgenom-Assembly CanFam3.1 gemappt. Das variant calling von Einzelnukleotid- und kleinen Indel-Varianten (SNV) wurde durchgeführt, und die NCBI-Annotation (Release 105) wurde zur Vorhersage ihrer funktionellen Auswirkungen wie zuvor beschrieben verwendet [24]. DNA-Varianten die ausschliesslich nur bei Leonbergern auftreten, die von allen Leonbergern geteilt werden, und seltene Varianten, die bei mindestens einem Leonberger auftreten, wurden durch Vergleich mit dem Variantenkatalog von 605 öffentlich zugänglichen Kontrollhunden aus 128 verschiedenen Rassen und 9 Wölfen identifiziert, der vom Dog Biomedical Variant Database Consortium [24] zur Verfügung gestellt wurde. Eine Liste von 113 Neuropathie- und Charcot-Marie-Tooth-Krankheit-assoziierten

Genen wurde aus den OMIM [41] und OMIA [42] Datenbanken extrahiert (Zusatzdatei 2). Diese Gene wurden für die Filterung aller 653 verfügbaren Genome auf solche Varianten die gehäuft in den sequenzierten Leonberger Hunden vorkommen verwendet und die Allelhäufigkeit in den übrigen Hunderassen wurde berechnet, um nach seltenen populationsspezifischen krankheitsassoziierten Varianten zu suchen, die möglicherweise von anderen mit dem Leonberger verwandten Rassen geteilt werden. Mehrere in-silico Vorhersagewerkzeuge, PROVEAN [43], MutPred2 [44], MutPred-Indel [45], MutPred-LOF [46] und PredictSNP [47], wurden verwendet, um die möglichen Effekte der entdeckten Varianten auf die entsprechenden Proteine vorherzusagen. Die Variabilität des mitochondrialen Genoms wurde untersucht, um die Diversität in den Haplogruppen der untersuchten Hunde unter Verwendung der Nomenklatur zu bestimmen, die sowohl von Pereira et al. [48] als auch von Duleba et al. [29] beschrieben wurde. Der Integrative Genomics Viewer wurde zur visuellen Inspektion und Bestätigung der detektierten SNVs verwendet [49].

Resultate

Stammbaum-Analysen

Es standen Stammbauminformationen von 142.072 Leonbergern zur Verfügung, wobei der Index der Stammbaumvollständigkeit über 5 Generationen bei Tieren aus den jüngsten Kohorten über 99 % und im Jahr 1935 über 80 % lag. Seit dem Jahr 2000 wurden weltweit jedes Jahr mehr als 4.000 Hunde geboren (Zusatzdatei 3). Die durchschnittliche Wurfgröße über die Kohorten (*Gruppe mit gemeinsamen zeitspezifischen demografischen Merkmalen*) hinweg betrug 6,5 Welpen und es wurde ein konstantes Generationsintervall von 4 Jahren beobachtet. **Insgesamt wurden 22 Gründertiere identifiziert (10 Rüden, 12 Hündinnen), von denen 3 (2 Rüden, 1 Hündin) zusammen fast 46 % zur letzten Kohorte beitrugen (Geburtsjahr 2016).** Die Leonberger-Rasse erlebte in den 1940er Jahren einen starken Bottleneck-Effekt, als 1946 nur 17 Hunde ($F_{\text{PED}}=[0.14-0.35]$) registriert wurden (Zusatzdatei 3).

Der durchschnittliche stammbaumbasierte Inzuchtkoeffizient (F_{PED}) zeigte zusammen mit der durchschnittlichen Verwandtschaft einen stetigen Anstieg über die analysierten Jahre, wobei F_{PED} in der letzten Kohorte (Geburtsjahrgang 2016) auf 0,29 geschätzt wurde (Abbildung 1A). Ein beliebter Vater wurde definiert als ein Rüde, der mindestens 33 Welpen gezeugt hat, was 5 Würfen auf der Grundlage der beobachteten durchschnittlichen Wurfgröße entspricht. Für 5.456 Vererber, die zwischen 1894 und 2016 geboren wurden, waren Informationen über die Anzahl der Nachkommen verfügbar. **Ein populärer Vätereffekt war offensichtlich, da ein Viertel aller Väter zwei Drittel aller Nachkommen produzierte** (Abbildung 1B). Zusätzlich, obwohl der 1985 geborene Top-Zuchtrüde 434 Hunde gezeugt hat, haben der zweite und dritte Top-Zuchtrüde, beide 2003 geboren, 394 bzw. 337 registrierte Welpen gezeugt. Dieser Trend ist in der gesamten Zuchtgeschichte des Leonbergers sichtbar (Zusätzliche Datei 3): die Top 27 Vererber, die zwischen 1976 und 2010 geboren wurden, haben jeweils mehr als 200 Welpen produziert (was mindestens 30 Würfen pro Vererber entspricht). Zusätzlich war das Todesdatum von 4.783 Hunden (2.453 Hündinnen und 2.330 Rüden) bekannt, die zwischen 1959 und 2019 geboren wurden. Die Lebenserwartung wurde weiterhin für die Jahre 1989-2004 ausgewertet, in denen mehr als 100 Datensätze verfügbar waren und davon ausgegangen werden konnte, dass es keine lebenden Hunde mehr gab. Diese Kohorte umfasste 3.044 Hunde (1460 Rüden und 1584 Hündinnen) und ergab eine mittlere Lebenserwartung von 8,2 (Median 8,5) Jahren, wobei die Lebenserwartung der Rüden (Mittelwert=7,9) niedriger war als die der Hündinnen (Mittelwert=8,5). Darüber hinaus wurde ein langsamer, kontinuierlicher Rückgang über die Zeit beobachtet, von durchschnittlich 9,4 Jahren im Jahr 1989 auf 7,7 Jahre im Jahr 2004 (Median: 10,1 bis 8,2 Jahre) (Abbildung 1C).

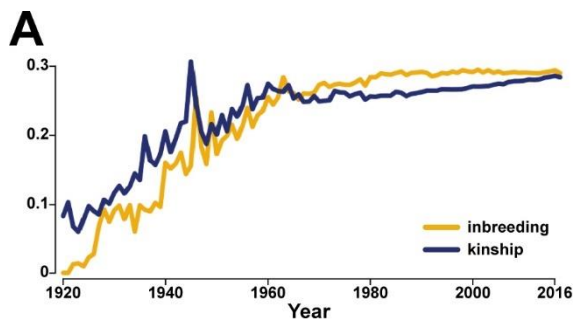
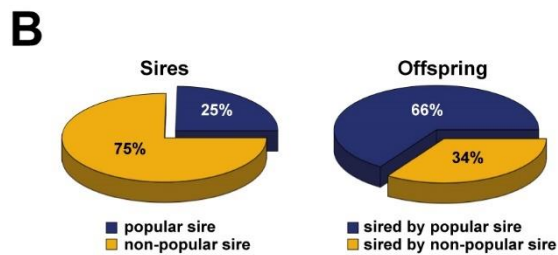
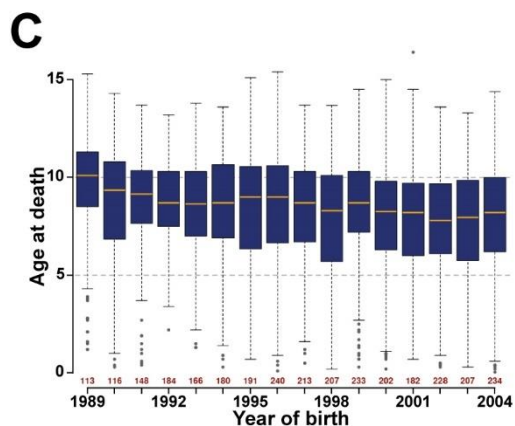


Abbildung 1 Stammbaumanalysen von Leonberger Hunden.

A. Durchschnittlicher stammbaumbasierter Inzuchtkoeffizient und geschätzte Verwandtschaft (MK) pro Jahr von 1920 bis 2016.



B. Tortendiagramme des Anteils der in der Zucht verwendeten Vererber zeigen deutlich das beliebte Vätereffekt. Beliebte Vererber sind Rüden, die mindestens 33 Welpen produziert haben und 25 % aller Zuchtrüden ausmachen. Diese Vererber sind für 66 % aller Nachkommen verantwortlich.



C. Boxplot des Sterbealters pro Geburtsjahr, der die Lebenserwartung der Leonberger Hunde über die Jahre 1989-2004 zeigt. Die genaue Anzahl der verfügbaren Datensätze pro Jahr ist in Rot über jedem Jahr angegeben.

Die mittlere Verwandtschaft wurde an einem Datensatz von 31.832 Leonberger Hunden analysiert, die potentiell für die Zucht zur Verfügung eingesetzt werden könnten, was zu einer durchschnittlichen Verwandtschaft von 0,31 führte. Der Unterschied zwischen den minimalen (0,29) und maximalen (0,33) Werten der MK war gering, und es wurden keine Linien gefunden, die in hohem Maße nicht mit der Gesamtpopulation verwandt wären.

Zur besseren Visualisierung wurden drei Gruppen mit den Farben grün ($MK < 0,31$), gelb ($MK = [0,31 - 0,32]$) und orange ($MK > 0,32$) gebildet. Der Anteil der Hunde mit höherer oder niedrigerer MK unterschied sich in den verschiedenen Ländern (Zusätzliche Datei 4). Die farbkodierten MK-Koeffizienten sind in die Worldwide Independent Leonberger Database [2] eingeflossen.

Leonberger Hunde sind anfällig für mehrere Erkrankungen

Wir sammelten medizinische Informationen über die gesamte Lebensdauer von 2.726 Leonbergern aus den von den Besitzern eingereichten Gesundheitsberichten. Von diesen litten 1.334 (48,9 %) Hunde an mindestens einer Beeinträchtigung des Gesundheitszustandes (Zusatzdatei 5). Einzelne Hunde ($n=544$) litten jedoch oft an mehreren Störungen. Bei insgesamt 586 (21,5 %) Hunden wurde ein Tumor oder Krebs festgestellt, wobei das Osteosarkom (42,5 %) und das Hämangiosarkom (22,5 %) die beiden häufigsten Arten waren. Andere häufig auftauchende Gesundheitsprobleme waren orthopädischer (15,8 %), neurologischer (14,8 %), endokrinologischer (5,6 %), verdauungsbedingter (4,2 %) und kardialer (4,1 %) Natur. Die häufigsten spezifischen Störungen in jeder Systemkategorie waren Arthritis (222 Fälle), Polyneuropathie (362 Fälle), Hypothyreose (142 Fälle), Magentorsion (54 Fälle) und dilatative Kardiomyopathie (45 Fälle). Alle gemeldeten Erkrankungen und die Anzahl der Hunde, die daran leiden, sind in der Zusatzdatei 5 aufgeführt. Die meisten von Polyneuropathie betroffenen Hunde (71,0 %) zeigten sowohl Atem- als auch Ganganomalien, während 14,6 % nur eine Larynxlähmung oder Atemprobleme und 14,4 % nur Ganganomalien aufwiesen (Zusatzdatei 5).

MDS-Analyse der weltweiten Leonberger-Populationsstruktur

Die 1.203 Hunde, die auf SNP-Arrays genotypisiert wurden, repräsentierten die Leonberger-Population gut, da sie im Laufe der letzten 10 Jahre und aus verschiedenen Ländern kamen (Zusatzdatei 1). Aufgrund der großen geografischen Entfernungen zwischen den Hunden wurden Subpopulationen erwartet: 579 Proben (48,1 %) kamen aus den USA und Kanada, 602 Proben (50,0 %) kamen aus europäischen Ländern und 22 Proben (1,8 %) kamen aus Australien, Neuseeland oder Japan. Die multidimensionale Skalierung (MDS) der paarweisen genetischen Abstände zeigte jedoch keine offensichtliche Clusterung (*Gruppenbildung*) (Figure 2A, Zusatzdatei 6), obwohl eine kleine Gruppe nordamerikanischer Hunde sich nicht vollständig mit dem Rest der Population überschneidet. Wie sich in Zusatzdatei 6B zeigt, wurde die sich nicht überlappende Hundegruppe mehrheitlich vor dem Jahr 2000 geboren, während die erst kürzlich geborenen Hunde näher an der europäischen Population zu sein scheinen. Zusätzlich muss der offensichtliche Unterschied in der Anzahl der genotypisierten Hunde, die im oder vor dem Jahr 2000 geboren wurden, aus Nordamerika (199 von 579) und andererseits aus Europa (88 von 602) berücksichtigt werden; diese Disproportion (Missverhältnis) in der Probenmenge könnte zu der nicht vollständigen Überlappung beitragen.

Analyse der runs of homozygosity

Der ROH-basierte genomische Inzuchtkoeffizient (F_{ROH}) lag im Durchschnitt bei 0,28 (mit Bereich [0,05-0,47]). **Darüber hinaus deckte der Vergleich von F_{PED} (Pedegree, Stammbaum) und F_{ROH} einige Hunde mit offensichtlich falschen Stammbaumeinträgen auf und ermöglichte die Schätzung der Inzucht bei Individuen mit fehlenden oder unbekanntem Stammbäumen** (Abbildung 2B, Zusatzdatei 1). Insgesamt wurden 125.653 ROHs mit einer durchschnittlichen Länge von 5,88 Mb (Bereich von [1,00-90,17]) identifiziert. ROHs konnten aufgrund der unterschiedlichen Größen in fünf Gruppen eingeteilt werden. Die meisten der entdeckten ROHs (53,6 %) waren kleiner als 4 Mb und 16,1 % waren größer als 10 Mb (Abbildung 2C). Die Anzahl der ROHs pro Chromosom reichte von 1.419 auf Chromosom 38 bis 5.303 auf Chromosom 1; der Anteil der gesamten chromosomalen Beteiligung in homozygoten Regionen im Verhältnis zu seiner chromosomalen Größe war am höchsten bei Chromosom 33 (16,9 %) und am niedrigsten bei Chromosom 1 (6,6 %) (Abbildung 2D). Im Durchschnitt wurden 104,4 ROHs pro Hund festgestellt, wobei die Spanne von 37 bis 177 reichte. Die Gesamtlänge des ROH über das Genom pro Individuum betrug im Durchschnitt 614 Mb und reichte von 104,6 bis 1028,2 Mb. Der Anteil der Hunde mit einem SNP innerhalb eines ROH variierte über die 38 Autosomen, wie in Zusatzdatei 7 gezeigt. Auf Chromosom 10 (chr10:44.529.255-44.831.343) teilten 1.047 Hunde (87,0 %) einen ROH über ~302kb. Nur ein Protein-kodierendes Gen ist in dieser Region annotiert, TMEM131, das für ein Transmembranprotein kodiert, von dem nicht bekannt ist, dass es mit einer Krankheit in Verbindung steht. Seine vermutete Funktion liegt in der Entwicklung und Differenzierung von T-Zellen sowie in der Kollagenrekrutierung und -sekretion [50, 51]. Eine ~236kb große Region (chr26:35.428.129-35.663.964) am Ende von Chromosom 26 wurde von 1.000 (83,1 %) Hunden geteilt (Abbildung 2E). Allerdings sind in dieser Region nur zwei Pseudogene und eine uncharakterisierte ncRNA annotiert, von denen BLAST-basierte Vergleiche keine signifikante Sequenzähnlichkeit zum Menschen ergaben.

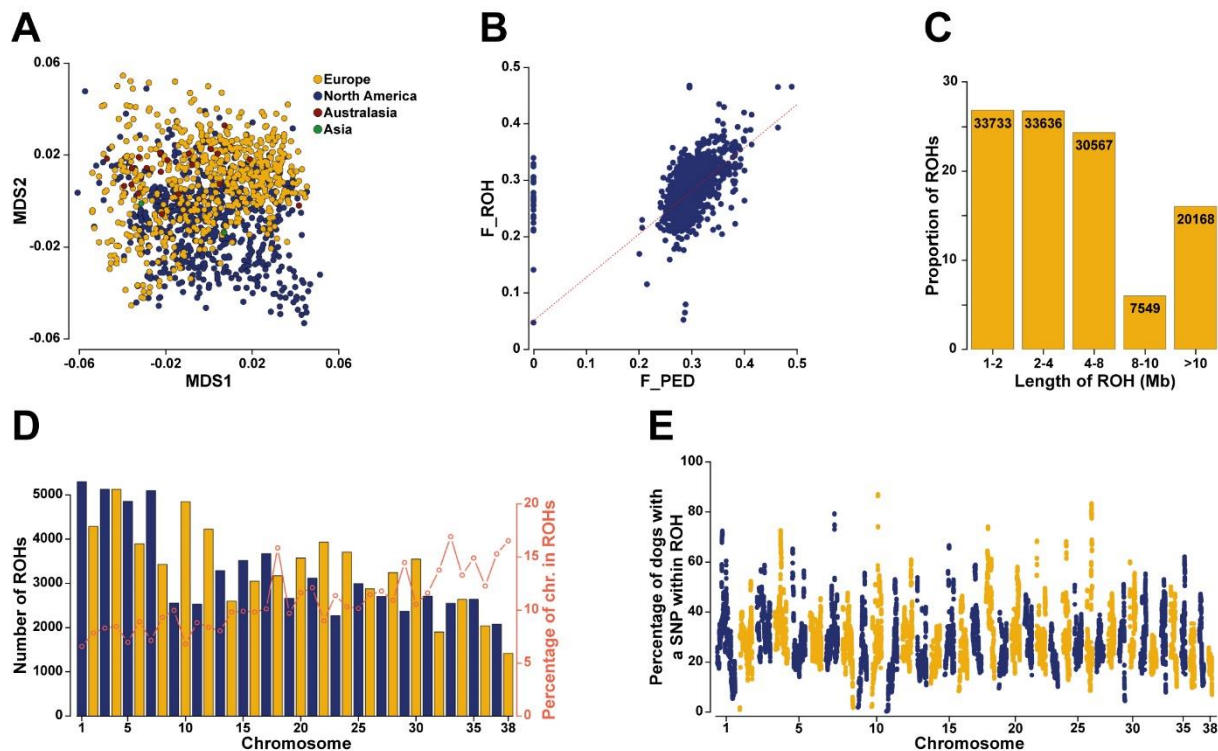


Abbildung 2 Grafische Darstellung der genomischen Analysen von Leonberger Hunden.
 A. Das MDS-Diagramm von 1203 Hunden zeigt unabhängig von der geografischen Region des Wohnortes keine offensichtlichen Subpopulationen.
 B. Korrelation der Inzuchtkoeffizienten, die aus den Stammbaumdaten (F_PED) und der ROH-Analyse (F_ROH) geschätzt wurden. Beachten Sie, dass die Hunde mit unbekanntem Stammbaum (F_PED=0) auf der linken Seite meist den gleichen Grad an Inzucht aufweisen wie die Hunde mit bekannten registrierten Aufzeichnungen.
 C. Häufigkeiten und Anzahl der verschiedenen Klassen von ROHs, basierend auf ihrer Länge, die sowohl alte als auch neue Inzucht anzeigen.
 D. Anzahl der erkannten ROHs pro Autosom (linke y-Achse) und der Anteil der Chromosomen in einem ROH (rechte y-Achse).
 E. Manhattan-Diagramm, das den Anteil der Hunde zeigt, die einen SNP innerhalb eines ROH über die 38 Autosomen teilen. Die meisten Hunde teilen eine Region auf Chromosom 10 und 26.

Whole genome sequence-Analyse (WGS) identifiziert keine fixierten proteinverändernden Varianten bei Leonbergern

WGS-Daten von 39 Leonberger-Hunden, darunter 38 Individuen mit verschiedenen neurologischen Symptomen (Zusatzdatei 1), wurden mit 614 öffentlich verfügbaren WGS-Daten anderer Hunderassen verglichen [24]. Zu 4 Genomen (WGS-Datensätze) von Leonbergern gehörten auch jene, die sequenziert wurden, um die zuvor beschriebenen pathogenen Varianten in den Genen GJA9 und NAPEPLD zu entschlüsseln [5, 6]. Wir suchten zunächst nach für Leonberger rassenspezifischen Varianten, die in den Kontrollgenomen fehlten, und identifizierten nur 27 nicht-kodierende und 1 kodierende populationsspezifische Variante, die von allen 39 Leonberger-Hunden in mindestens heterozygotem Zustand geteilt wurden (Zusatzdatei 8). Neun dieser Varianten befanden sich in unmittelbarer Nähe der detektierten intergenen ROH-Region auf Chromosom 26. Die einzige kodierende Variante war eine synonyme SNV im ZNF510-Gen, das ein Zinkfingerprotein kodiert.

Leonberger-spezifische proteinverändernde Varianten als neue Kandidaten für neuronale Erkrankungen

Als nächstes suchten wir nach seltenen Varianten (Allele, die in mindestens einem Leonberger vorkommen und in den 614 Kontrollgenomen fehlen), was zur Entdeckung von 62.531 SNVs führte, von denen 1379 als kodierend vorhergesagt wurden (Zusatzdatei 9). Von diesen wurden 855 als proteinverändernd vorhergesagt, wobei die Häufigkeit dieser Leonberger-spezifischen Allele

zwischen 0,01 und 0,63 lag (Zusatzdatei 9). Unter Berücksichtigung der Gene, die mit vererbten neuronalen Erkrankungen in anderen Spezies assoziiert sind, gab es 11 potentiell pathogene Varianten (Tabelle 1). Diese kurze Liste enthielt zwei bereits bekannte pathogene Varianten, nämlich die Leukoencephalomyelopathie-assoziierte Missense-Variante in NAPEPLD [6] mit einer Frequenz von 0,12 und die Polyneuropathie-assoziierte Frameshift-Variante in GJA9 [5] mit einer Frequenz von 0,05. Neben einer einzelnen Missense-Variante (*Punktmutation*), die als neutral vorhergesagt wurde, wurden 8 der weiteren 9 Varianten als wahrscheinlich schädlich vorhergesagt, darunter 2 Missense-, 2 Frameshift-, 2 In-Frame-Insertionen, 1 In-Frame-Deletion und 1 Nonsense-Variante (Tabelle 1, Zusatzdatei 9, Zusatzdatei 10).

Tabelle 1 Seltene proteinverändernde Varianten einzig bei Leonberger Hunden

Gene	OMIM/ OMIA number	Variant designation ¹			Alternative allele frequency
		Genomic position	Coding DNA change	Protein change	
<i>ATXN7</i>	607640	chr20:27234549	c.2285T>C	p.Val762Ala	0.1154
<i>CEP55</i>	610000	chr28:7777489	c.1334A>C	p.Tyr446Ser	0.0256
<i>CNTF</i>	118945	chr18:37758771	c.401_402insA	p.Asn137fs	0.0128
<i>ELOVL4</i>	605512	chr12:40850011	c.424_433delGGAG CACAGC	p.Gly142fs	0.0128
<i>FDXIL</i>	614585	chr20:50798661	c.124_125insGGCC GCCATCACCACG GCGGTGAGCACC GCCGCCAGCAGC ACCAGCCCGTCA GCGTTGAGCCG	p.Thr41_Ala4 2insGlyProPro SerProArgArg Ter	0.0128
<i>GJA9</i> ²	611923/2 119-9615	chr15:3863519	c.1108_1109delGA	p.Glu370fs	0.0513
<i>MCM3AP</i>	603294	chr31:39553337	c.1785_1793delCTC TGAAGG	p.Ser596_Gly5 98del	0.0128
<i>NAPEPLD</i> ³	612334/1 788-9615	chr18:16987520	c.559G>C	p.Ala187Pro	0.1154
<i>PLEKHG5</i>	611101	chr5:60325903	c.1585C>T	p.Gln529*	0.0128
<i>SPTBN4</i>	606214/2 232-9823	chr1:113215064	c.1247_1248insGGT AGCCCATGCGGT	p.Ala416_Ala 417insValAla HisAlaVal	0.0128
<i>SYNE1</i>	608441	chr1:42549994	c.17359C>T	p.Arg5788Trp	0.0256

¹Additional details including the protein prediction effects are described in Additional file 10. All positions refer to the CanFam3.1 reference sequence assembly.

²Previously described polyneuropathy-associated variant [5].

³Previously described leukoencephalomyelopathy-associated variant [6].

Anreicherung von Proteinverändernden Varianten in Kandidatengenen für neuronale Störungen bei Leonbergern

Da 33 der 39 sequenzierten Leonberger an einer ungeklärten Form der Polyneuropathie litten (Zusatzdatei 1), wurde zusätzlich nach Varianten gesucht, die in 113 mutmasslichen Kandidatengenen (Zusatzdatei 2) vorkommen, die an der Entstehung der unerklärlichen Form der Polyneuropathie beteiligt sein könnten. Die Fokussierung auf kodierende Varianten, die in mindestens einem sequenzierten Leonberger vorkommen, ergab 649 Varianten, die 100 Gene betreffen, von denen 232 Varianten als proteinverändernd vorhergesagt wurden (Zusatzdatei 11). Alle diese Varianten waren mit variabler Häufigkeit auch bei anderen Hunderassen sowie bei Wölfen vorhanden.

Um nach Varianten zu suchen, die beim Leonberger gehäuft vorkommen, betrachteten wir als nächstes solche mit einer Allelhäufigkeit >0% und <10% bei den 614 Kontrollhunden und mit einer

mehr als doppelt so hohen Allelhäufigkeit bei den Leonberger Hunden gegenüber den Kontrolltieren. Dieser Ansatz ergab eine kurze Liste von 22 SNVs, die 17 Gene betreffen, darunter 21 Missense- und 1 In-Frame-Deletions-Variante (Tabelle 2). Neben den 11 Varianten, die als neutral vorausgesagt wurden, wurden 11 Varianten als höchstwahrscheinlich schädlich vorausgesagt, was potentiell pathogene Varianten darstellen, die Polyneuropathie verursachen (Tabelle 2, Zusatzdatei 10). Interessanterweise wurde eine neue potenziell schädliche Missense-Variante an einem anderen Codon im GJA9-Gen gefunden, in dem sich eine frühere Variante als polyneuropathie-assoziiert bei Leonbergern erwiesen hat [5]. Diese neu entdeckte GJA9-Variante trat mit einer recht hohen Häufigkeit (0,55) innerhalb der sequenzierten Kohorte von Leonberger Hunden auf, war aber auch vergleichsweise häufig (0,09) unter den Kontrollen aus 35 verschiedenen Rassen (Zusatzdatei 9). Die anderen Varianten traten bei verschiedenen Rassen mit unterschiedlicher Häufigkeit auf (Zusatzdatei 9). Nur eine Missense-Variante im DST-Gen war recht selten und war nur bei einem einzigen Howavart-Hund in einem heterozygoten Zustand vorhanden (Zusatzdatei 9).

Tabelle 2 Proteinverändernde Varianten in Polyneuropathie-assoziierten Kandidatengeneten, angereichert bei Leonberger Hunden im Vergleich zu 614 Kontrollen verschiedener Rassen

Gene	OMIM/ OMIA number	Variant designation ¹			Alternative allele frequency	
		Genomic position	Coding DNA change	Protein change	Leonbergers	controls
<i>CNTNAP1</i>	602346	chr9:20294320	c.3863G>C	p.Arg1288Pro	0.3974	0.0366
<i>CNTNAP1</i>	602346	chr9:20298261	c.2585G>A	p.Gly862Glu	0.1538	0.0067
<i>DHTKD1</i>	614984	chr2:24316951	c.1820C>T	p.Ala607Val	0.0897	0.0049
<i>DIAPH3</i>	614567	chr22:15880854	c.2000G>A	p.Cys667Tyr	0.1538	0.0689
<i>DST</i>	113810	chr12:23771235	c.18224C>T	p.Thr6075Met	0.0769	0.0008
<i>DST</i>	113810	chr12:23782332	c.17800C>T	p.Arg5934Trp	0.4359	0.0330
<i>DST</i>	113810	chr12:23846624	c.10661C>T	p.Ser3554Leu	0.0769	0.0369
<i>DYNC1H1</i>	600112	chr8:70064306	c.13757C>T	p.Pro4586Leu	0.1282	0.0051
<i>GARS</i>	600287	chr14:43322005	c.622G>A	p.Val208Ile	0.1154	0.0025
<i>GJA9</i>	611923/2 119-9615	chr15:3862761	c.344G>C	p.Arg115Thr	0.5513	0.0868
<i>JPH1</i>	605266	chr29:22710800	c.1531A>G	p.Ile511Val	0.1923	0.0299
<i>LOC477508</i>	602072	chr26:16348822	c.298G>A	p.Val100Met	0.1795	0.0066
<i>MME</i>	120520	chr23:49045461	c.1700A>T	p.Gln567Leu	0.0769	0.0157
<i>NEFH</i>	162230	chr26:22729485	c.1046G>A	p.Arg349His	0.2821	0.0116
<i>NEFH</i>	162230	chr26:22732974	c.1901_1924 delTGAAGG AGGAGGC CAAGTCCC CAG	p.Val640_Pro6 47del	0.3974	0.0849
<i>NEFH</i>	162230	chr26:22733415	c.2326C>G	p.Pro768Ala	0.0513	0.0059
<i>OTOF</i>	603681	chr17:20534866	c.3451G>A	p.Ala1151Thr	0.7051	0.0094
<i>PDXK</i>	179020	chr31:37707445	c.774G>C	p.Arg258Ser	0.1923	0.0898
<i>PRX</i>	605725	chr1:113290407	c.784G>A	p.Ala262Thr	0.1282	0.0257
<i>SBF2</i>	607697	chr21:33036988	c.4114C>T	p.Pro1372Ser	0.1667	0.0361
<i>SLC12A6</i>	604878	chr30:853540	c.2498C>T	p.Ala833Val	0.0513	0.0250
<i>WNK1</i>	605232	chr27:42911057	c.7448C>G	p.Thr2483Arg	0.0128	0.0058

¹Additional details including the protein prediction effects are described in Additional file 10.

All positions refer to the CanFam3.1 reference sequence assembly.

Mitochondriale Genom-Diversität ist bei Leonbergern begrenzt

Schließlich wurde die Diversität des mitochondrialen Genoms der 39 untersuchten Leonberger ausgewertet und ergab nur zwei Haplotypen. Betrachtet man nur den mtDNA D-Loop Nomenklatur-Standard [48], so zeigten 38 Hunde den A2-Haplotyp und ein einzelner Hund hatte den A17-Haplotyp. Die gleichen Ergebnisse wurden bei der Verwendung der vollständigen Mitogenom-Datenbank mit standardisierter Haplogruppen-Nomenklatur [29] erzielt, wobei 38 Hunde zur Haplogruppe A1b2a1a1 und der eine Hund zur Haplogruppe A1b1a1a gehörten (Zusatzdatei 12). Zusätzlich wurden insgesamt 10 weitere SNVs im kodierenden Teil der mtDNA nachgewiesen, davon 7 die ausschliesslich bei einzelnen Leonbergern und in der Rasse insgesamt selten vorkommen und bei den 614 Kontrollen nicht gesehen wurden (Zusatzdatei 12). Diese Liste enthielt 2 Varianten, die tRNA-Gene betreffen, und eine einzelne Missense-Variante im ND2-Gen, die bei einem einzigen Hund beobachtet wurde (Tabelle 3).

Tabelle 3 Varianten in der mitochondrialen DNA von Leonberger Hunden

Gene	Variant designation¹	Proportion of variant²	OMIM/OMIA number	Type
<i>tRNA-Phe</i>	m.49A>G	96%	590070	unknown
<i>tRNA-Val</i>	m.1069T>C	99%	590105	unknown
<i>ND2</i>	m.4764T>C; p.Ile284Thr	100%	516001	missense
<i>COX1</i>	m.5681A>G	66%, 89%	516030	synonymous
<i>ND4</i>	m.11211C>T	72%, 66%	516003	synonymous
<i>ND5</i>	m.13118T>C	62%	516005	synonymous
<i>ND5</i>	m.13544A>G	100%, 100%, 100%, 100%	516005	synonymous

¹All positions refer to the CanFam3.1 reference sequence assembly.

²The percentage of the variant represents heteroplasmy, 100% represents homoplasmy, for each dog that carries the variant calculated from the coverage of the affected mtDNA region.

Diskussion

Dies ist die erste Studie, die die genomische Diversität und Populationsstruktur der Rasse Leonberger umfassend beschreibt, indem sie die ROH-Verteilung und SNVs anhand von SNP-Array- und WGS-Daten untersucht. Darüber hinaus wurden die genealogischen Stammbauminformationen verwendet, um die geschätzten Inzuchtkoeffizienten und die mittlere Verwandtschaft in der aktuellen Population zu vergleichen. Den neu entdeckten genomischen Regionen könnten rassespezifische Merkmale zugrunde liegen und sie könnten Informationen über die genetische Struktur der Rasse liefern, wie z. B. den Selektionsdruck auf bestimmte Merkmale, Inzuchtgrade oder genetische Engpässe.

Stammbaumanalysen, die auf umfangreichen Daten basieren, ergaben ähnliche Zahlen zur durchschnittlichen Wurfgröße und Lebenserwartung für Leonberger Hunde, die mit den Ergebnissen einer früheren Studie übereinstimmen [52]. Während längere ROH-Segmente (~10 Mb) als Ergebnis Inzucht der jüngeren Zeit auftreten, zeigen kürzere ROHs (<2 Mb) genomische Regionen an, die kennzeichnend für die Spuren der bereits länger bestehenden Inzucht sind, diese abstammungsidentischen Genomabschnitte wurden durch Rekombinationsereignisse während zahlreicher Generationen deutlich verkürzt [12]. Bei den untersuchten Leonberger-Hunden machen die 1-2 Mb großen Segmente etwa 25 % der entdeckten ROHs aus, während etwa 20 % der Segmente länger als 8 Mb sind (Abbildung 2C), was sowohl auf frühere als auch auf neue Inzucht hinweist. Dies steht im Einklang mit der bekannten Geschichte der Rasse und den beobachteten aktuellen Zuchtpraktiken. Kleinere ROHs sind aufgrund der begrenzten Anzahl von Markern und der unterschiedlichen Abstände im Genom schwer zu erkennen. Zum Beispiel wurde das stark geteilte

ROH auf Chromosom 26, für das 83 % der Hunde homozygot waren, durch die Identifizierung von 9 Varianten, die alle 39 sequenzierten Hunde gemeinsam hatten, bestätigt. Die Stammbaumanalyse zeigte einen klaren Bottleneck Effekt während der 1940er Jahre, sowie die intensive Nutzung von populären Vererbern und der Linienzucht, was zu dem jüngsten hohen Verwandtschaftsgrad unter den Leonbergern führte. Eine begrenzte Anzahl von Vorfahren mit einem hohem genetischen Beitrag hat einen großen Einfluss auf die Verwandtschaft innerhalb der Rasse, was zu Inzucht aufgrund dieser Verwandtschaft und damit zur Inzuchtdepression führt. Mögliche krankheitsverursachende genetische Varianten, die diese Vorfahren trugen, sind nun wahrscheinlich in der gesamten Population verbreitet. Darüber hinaus zeigte die MDS-Analyse keine klare geografische oder abstammungsbedingte Differenzierung in der Leonberger-Population, im Gegensatz zu anderen Rassen von denen früher berichtet wurde, wie z.B. den Englischen Windhunden und Labrador Retrievern, die sich auf der Basis spezifischer Arbeits- und Showlinien unterschiedliche Gruppen bilden; oder die italienischen Windhunde und Shetland-Schäferhunde, bei denen die europäischen und nordamerikanischen Populationen getrennte Cluster bilden [53]. Unsere weltweite Gesundheitsbefragung von fast 3.000 Leonbergern bestätigte eine hohe Prävalenz von Krebserkrankungen, insbesondere Osteosarkomen und Hämangiosarkomen, neuromuskulären Störungen und Hypothyreose. Bei anderen Hunderassen haben genetische Assoziationsstudien zur Identifizierung genetischer Risikofaktoren [27] eine starke Assoziation mit bestimmten Genomregionen ergeben; z. B. für Osteosarkome [54, 55]. Ähnliche Studien könnten beim Leonberger durchgeführt werden, um Verknüpfungen und Marker für die Selektion gegen das erhöhte Risiko bestimmte Krebsformen zu entwickeln, zu entschlüsseln. Auch die Hypothyreose wurde recht häufig festgestellt, was frühere Berichte über diese Rasse bestätigen [56, 57]. Dagegen traten Erkrankungen wie Herzrhythmusstörungen, die zum plötzlichen Herztod führen [58] und von denen berichtet wurde, dass sie bei Leonbergern gehäuft vorkommen, in unserer Untersuchung nur sporadisch auf. Neben den orthopädischen Erkrankungen, von denen bekannt ist, dass sie polygenetischer Natur sind, waren neurologische Erkrankungen die zweithäufigste Krankheitskategorie beim Leonberger, mit überwiegend verschiedenen Formen der Polyneuropathie, wie bereits berichtet [59]. Eine kürzlich von der Leonberger Health Foundation International durchgeführte Studie, die auf mehr als 1.000 Hunden aus 24 verschiedenen Ländern basierte, wies auch auf die hohe Prävalenz bestimmter Krebsformen wie Osteosarkom und Hämangiosarkom sowie neurologischer Störungen wie Kehlkopflähmung und andere Formen der Polyneuropathie hin [60]. Obwohl diese beiden Erhebungen nicht wirklich umfassend und daher möglicherweise nicht vollkommen repräsentativ waren, wurden sie unabhängig voneinander zu unterschiedlichen Zeitpunkten durchgeführt, weshalb wir sie als angemessene Schätzungen der tatsächlichen Häufigkeiten betrachteten.

Wir sequenzierten die Genome von mehr als 30 an Polyneuropathie erkrankten Hunden mit dem Ziel, weitere krankheitsverursachende Varianten zu entschlüsseln. Die Genomsequenzdaten lieferten Hinweise auf sowohl rassespezifische als auch Genvarianten die vermehrt beim Leonberger (im Vergleich zu allen anderen Hunderassen) vorkommen, in den Genomen von an Polyneuropathie erkrankten Leonbergern mit unterschiedlichen Häufigkeiten. Die kurze Liste von 21 potentiell pathogenen Varianten in Kandidatengenomen für neurologische Erkrankungen enthält die beiden zuvor berichteten Varianten in GJA9, die Polyneuropathie verursachen [5], und in NAPEPLD, die Leukoenzephalomyelopathie verursachen [6], und bestätigt damit die potentielle Nützlichkeit des gewählten Ansatzes. Die neu entdeckten Varianten sind bekannte Kandidaten für neurologische Erkrankungen bei verschiedenen Spezies. Zukünftige Genotypisierungen der 19 neu identifizierten, potenziell schädlichen, proteinverändernden Varianten in Kohorten von gut phänotypisierten Leonberger Hunden sind erforderlich, um eine Verknüpfung dieser Varianten mit der Krankheit zu bestätigen. Die zweite potenziell pathogene Missense-Variante in GJA9 ist überzeugend, da sie auch bei Hunden nicht verwandter Rassen in geringerer Häufigkeit vorhanden ist. Die zwei der drei mtDNA-SNVs, die tRNA-Gene betreffen, werden ebenfalls interessant zu verfolgen sein, da eine Variante in tRNA-Tyr eine bekannte Form der sensorischen ataktischen Neuropathie bei Golden Retrievern verursacht [61]. Da wir bisher nicht in der Lage waren, durch genomweite Assoziationsstudien zusätzliche Loci zu identifizieren, die mit einer mendelschen Vererbung (*ein Gen*

= ein Merkmal) der Polyneuropathie assoziiert sind, halten wir es für durchaus möglich, dass eine Untergruppe dieser Störung eine komplexe genetische Natur hat (*mehrere Gene, sowie Umweltfaktoren können für die Merkmalsausprägung verantwortlich sein*) und die individuelle Wahrscheinlichkeit, die Störung zu entwickeln, von einer Kombination mehrerer Allele an mehreren Loci zusätzlich von Umweltfaktoren abhängt.

Die Untersuchung des Vorhandenseins großer struktureller Variationen im gesamten Leonberger-Genom wäre in Zukunft ebenfalls aufschlussreich, wenn eine verbesserte kynologische Referenzassemblierung (*Ein Referenzgenom, auch als Referenzassemblierung bekannt, ist eine digitale Nukleinsäuresequenzdatenbank, die von Wissenschaftlern als repräsentatives Beispiel für den Satz von Genen in einem idealisierten einzelnen Organismus einer Spezies zusammengestellt wurde*) und Annotation, also der Bestimmung der genauer Lokalisation der Gene, zusätzlich zu Long-Read-Sequenzdaten verfügbar werden. Frühere Studien schlugen einen Geschlechterbias (*Geschlechterverzerrung*) vor, wobei mehr männliche Leonberger von Polyneuropathie betroffen waren als weibliche [62], bis jetzt wurde kein Hauptrisikofaktor auf dem Chromosom X identifiziert.

Schlußfolgerungen

Beim Leonberger sind durch einen genetischen Flaschenhalseffekt im letzten Jahrhundert, verbunden mit einer danach zunehmenden Populationsgröße, erhebliche Verluste der genetischen Vielfalt zu verzeichnen. Diese Situation scheint in erster Linie auf die bevorzugt stattfindende Verwendung beliebter Vatertiere zurückzuführen zu sein, die zu einem hohen Maß zur Inzucht führen, was auch die Verbreitung unerwünschter genetischer Merkmale innerhalb des Genpools begünstigt hat. Die Aufrechterhaltung des gegenwärtigen Niveaus der genetischen Vielfalt wird nur durch klare Selektionsentscheidungen möglich sein, einerseits durch den Einbezug von insgesamt deutlich mehr Hunden in die Zucht und insbesondere die Vermeidung von wiederholten Anpaarungen (Wurfwiederholungen) sowie der Begrenzung des Einsatzes beliebter Vererber, nebeneiner Minimierung der Verwandtschaft zwischen den ausgewählten Anpaarungspartnern. Die ermittelten mittleren Verwandtschaftsgrade (*MK*) können Züchtern, zusätzlich zu den bereits verwendeten stammbaumbasierten Inzuchtkoeffizienten, bei der Auswahl geeigneter Verpaarungen helfen. Darüber hinaus könnte eine sorgfältige Auskreuzung dazu beitragen, langfristig die genetische Vielfalt zu optimieren, insgesamt die Heterozygotie in der Population zu erhöhen und zudem die Häufigkeit von krankheitsverursachenden Allelen zu reduzieren, um so das Auftreten verschiedener Gesundheitsprobleme zukünftig zu verringern. Die hier vorgelegte Liste der mutmasslichen krankheitsverursachenden Varianten in Kandidatengenen für neurologische und neuromuskuläre Erkrankungen könnte in Zukunft die Identifizierung weiterer tatsächlich krankheitsverursachender Mutationen ermöglichen und somit helfen bislang ungeklärte Formen z.B. der Polyneuropathie aufzuklären. Sie veranschaulicht auch die höchstwahrscheinlich heterogene genetische Architektur dieser Gruppe von Krankheiten, von der man annahm, dass sie einer einfachen Vererbung nach Mendel folgt, aber offensichtlich insgesamt betrachtet mehr und mehr genetisch komplexer, polygener Natur ist.

Liste der Abkürzungen

BLAST: Basic local alignment search tool

F_PED: pedigree-based coefficient of inbreeding

F_ROH: genomic-based coefficient of inbreeding

MDS: multidimensional scaling

MK: mean kinship/mittlere Verwandtschaft

mtDNA: mitochondriale DNA

OMIA: Online Mendelian Inheritance in Animals

OMIM: Online Mendelian Inheritance in Man

ROH: run of homozygosity

SNP: single nucleotide polymorphism

SNV: single nucleotide variant

WGS: whole-genome sequencing

Danksagung

The authors are grateful to all Leonberger owners and breeders, the Leonberger Health Foundation, the Schweizerischer Leonberger Club, and the International Leonberger Union for their support during the entire project. Special thanks go to Wilma and Ben Kroon for sharing the Worldwide Independent Leonberger Database. We thank Nathalie Besuchet-Schmutz and Heidi Signer-Hasler for invaluable technical assistance. The Next Generation Sequencing Platform and the Interfaculty Bioinformatics Unit of the University of Bern are acknowledged for performing the WGS and providing high-performance computational infrastructure.

Theorie

ROH: Auf Englisch runs of homozygosity. Sind Regionen im Genom eines Organismus welche kontinuierliche homozygote Regionen aufweisen. Diese entstehen aufgrund von Eltern, die identische Haplotypen an ihre Nachkommen weitergeben. Diese Technik kann verwendet werden, um den genomischen Fußabdruck von Inzucht in Erhaltungsprogrammen zu identifizieren, da Organismen, die kürzlich Inzucht erfahren haben, lange Läufe von Homozygotie aufweisen.

Bottleneck-Event: auf Deutsch Flaschenhals-Event beschreibt eine starke Verringerung der Grösse einer Population. Dieses Ereignis kann unterschiedliche Gründe haben wie Hungersnöte, Brände, Krankheiten etc. Solche Ereignisse können die Variation im Genpool einer Population verringern; danach verbleibt eine kleinere Population mit einer geringeren genetischen Vielfalt, um die Gene durch sexuelle Fortpflanzung an zukünftige Generationen von Nachkommen weiterzugeben. Die genetische Vielfalt bleibt geringer und nimmt nur zu, wenn es zu einem Genfluss aus einer anderen Population kommt oder sie nimmt mit der Zeit sehr langsam zu, da zufällige Mutationen auftreten.

Allel: Ein Allel ist eine von zwei oder mehr Versionen des gleichen Gens an der gleichen Stelle auf einem Chromosom. Die meisten Allele führen zu keiner oder nur zu einer geringen beobachtbaren Veränderung in der Funktion des Proteins, für das das Gen kodiert. Manchmal können jedoch unterschiedliche Allele zu unterschiedlichen beobachtbaren phänotypischen Merkmalen führen. Wenn die beiden Chromosomen für ein bestimmtes Gen das gleiche Allel enthalten, sind sie und der Organismus homozygot in Bezug auf dieses Gen. Wenn die Allele unterschiedlich sind, sind sie und der Organismus in Bezug auf dieses Gen heterozygot.

WGS: Whole genome sequencing: Ganzgenomsequenzierung

List of abbreviations

BLAST: Basic local alignment search tool

F_PED: pedigree-based coefficient of inbreeding

F_ROH: genomic-based coefficient of inbreeding

MDS: multidimensional scaling

MK: mean kinship

mtDNA: mitochondrial DNA

OMIA: Online Mendelian Inheritance in Animals

OMIM: Online Mendelian Inheritance in Man

ROH: run of homozygosity

SNP: single nucleotide polymorphism

SNV: single nucleotide variant

WGS: whole-genome sequencing

Declarations

Ethics approval and consent to participate

Dog blood or cheek swab samples were obtained via elective owner submission for diagnostic purposes or were submitted for genotyping of the previously reported polyneuropathy-associated variants in Leonbergers. Written consent was obtained from all dogs' owners.

Consent for publication

Not applicable.

Availability of data and materials

The WGS dataset generated and analysed during the current study is available in the European Nucleotide Archive (<http://www.ebi.ac.uk/ena>). The 39 Leonberger dog genomes have been made freely available under study accession number PRJEB16012 and all sample accession numbers are listed in Additional file 1. The SNP genotyping data is available on reasonable request.

Competing interests

Both the University of Minnesota and the University of Bern offer genotyping tests for polyneuropathy- and leukoencephalomyelopathy-associated variants in their respective laboratories and all proceeds from these tests go toward ongoing canine genetic research. The authors declare that they have no other competing interests.

Funding

Not applicable.

Authors' contributions

AL carried out the analyses, visualized the results, and drafted the manuscript. KMM contributed data and critically revised the manuscript. VJ performed bioinformatics. FRS contributed to the pedigree analysis and interpretation of data. PO performed the mean kinship analysis and helped with the interpretation of data. JRM and CD designed the study and revised the manuscript. All authors participated in writing the manuscript and have read and approved the final version.

Acknowledgements

The authors are grateful to all Leonberger owners and breeders, the Leonberger Health Foundation, the Schweizerischer Leonberger Club, and the International Leonberger Union for their support during the entire project. Special thanks go to Wilma and Ben Kroon for sharing the Worldwide Independent Leonberger Database. We thank Nathalie Besuchet-Schmutz and Heidi Signer-Hasler for invaluable technical assistance. The Next Generation Sequencing Platform and the Interfaculty Bioinformatics Unit of the University of Bern are acknowledged for performing the WGS and providing high-performance computational infrastructure.

References

1. Bliss-Isberg C. The Leonberger: A Comprehensive Guide to the Lion King of Breeds. Revodana Publishing; 2016.
2. The Worldwide Independent Leonberger Database. 2005. <http://www.leonberger-database.com/>. Accessed 1 Jan 2019.
3. Gough A, Thomas A, O'Neill D. Breed predispositions to disease in dogs and cats: Third Edition. 2018.
4. Ekenstedt KJ, Becker D, Minor KM, Shelton GD, Patterson EE, Bley T, et al. An ARHGEF10 Deletion Is Highly Associated with a Juvenile-Onset Inherited Polyneuropathy in Leonberger and Saint Bernard Dogs. *PLoS Genet.* 2014;10:1–9.
5. Becker D, Minor KM, Letko A, Ekenstedt KJ, Jagannathan V, Leeb T, et al. A GJA9 frameshift variant is associated with polyneuropathy in Leonberger dogs. *BMC Genomics.* 2017;18:662.
6. Minor KM, Letko A, Becker D, Drögemüller M, Mandigers PJJ, Bellekom SR, et al. Canine NAPEPLD-associated models of human myelin disorders. *Sci Rep.* 2018;8:1–11.
7. Melis C, Borg ÅA, Espelien IS, Jensen H. Low neutral genetic variability in a specialist puffin hunter: The Norwegian Lundehund. *Anim Genet.* 2013;44:348–51.

8. Ontiveros ES, Hughes S, Penedo MCT, Grahn RA, Stern JA. Genetic heterogeneity and diversity of North American golden retrievers using a low density STR marker panel. *PLoS One*. 2019;14:1–15.
9. Yang Q, Chen H, Ye J, Liu C, Wei R, Chen C, et al. Genetic Diversity and Signatures of Selection in 15 Chinese Indigenous Dog Breeds Revealed by Genome-Wide SNPs. *Front Genet*. 2019;10:1–14.
10. Mortlock SA, Williamson P, Khatkar MS. Copy number variation and variant discovery in Bullmastiff dogs. *Anim Genet*. 2019;50:177–81.
11. Gajaweera C, Kang JM, Lee DH, Lee SH, Kim YK, Wijayananda HI, et al. Genetic diversity and population structure of the Sapsaree, a native Korean dog breed. *BMC Genet*. 2019;20:1–11.
12. Boccardo A, Marelli SP, Pravettoni D, Bagnato A, Busca GA, Strillacci MG. The German Shorthair Pointer Dog Breed (*Canis lupus familiaris*): Genomic Inbreeding and Variability. *Animals*. 2020;10:498.
13. Stronen A V., Salmela E, Baldursdottir BK, Berg P, Espelien IS, Jarvi K, et al. Genetic rescue of an endangered domestic animal through outcrossing with closely related breeds: A case study of the Norwegian Lundehund. *PLoS One*. 2017;12:1–18.
14. Curik I, Ferencaković M, Sölkner J. Inbreeding and runs of homozygosity: A possible solution to an old problem. *Livest Sci*. 2014;166:26–34.
15. McQuillan R, Leutenegger AL, Abdel-Rahman R, Franklin CS, Pericic M, Barac-Lauc L, et al. Runs of Homozygosity in European Populations. *Am J Hum Genet*. 2008;83:359–72.
16. Keller MC, Visscher PM, Goddard ME. Quantification of inbreeding due to distant ancestors and its detection using dense single nucleotide polymorphism data. *Genetics*. 2011;189:237–49.
17. Purfield DC, Berry DP, McParland S, Bradley DG. Runs of homozygosity and population history in cattle. *BMC Genet*. 2012;13:1.
18. Bai B, Zhao WM, Tang BX, Wang YQ, Wang L, Zhang Z, et al. DoGSD: The dog and Wolf genome SNP database. *Nucleic Acids Res*. 2015;43:D777–83.
19. Ostrander EA, Wayne RK, Freedman AH, Davis BW. Demographic history, selection and functional diversity of the canine genome. *Nat Rev Genet*. 2017;18:705–20.
20. Chu ET, Simpson MJ, Diehl K, Page RL, Sams AJ, Boyko AR. Inbreeding depression causes reduced fecundity in Golden Retrievers. *Mamm Genome*. 2019;30:166–72.
21. Metzger J, Distl O, Pfahler S. Variant detection and runs of homozygosity in next generation sequencing data elucidate the genetic background of Lundehund syndrome. *BMC Genomics*. 2016;17.
22. Kettunen A, Daverdin M, Helfjord T, Berg P. Cross-breeding is inevitable to conserve the highly inbred population of puffin hunter: The Norwegian Lundehund. *PLoS One*. 2017;12:1–16.
23. Bannasch D, Safra N, Young A, Karmi N, Schaible RS, Ling G V. Mutations in the *SLC2A9* gene cause hyperuricosuria and hyperuricemia in the dog. *PLoS Genet*. 2008;4:e1000246.
24. Jagannathan V, Drögemüller C, Leeb T, Aguirre G, André C, Bannasch D, et al. A comprehensive biomedical variant catalogue based on whole genome sequences of 582 dogs and eight wolves. *Anim Genet*. 2019;50:695–704.
25. Schoenebeck JJ, Ostrander EA. Insights into Morphology and Disease from the Dog Genome Project. *Annu Rev Cell Dev Biol*. 2014;30:535–60.
26. Lindblad-Toh K. What animals can teach us about evolution, the human genome, and human disease. *Ups J Med Sci*. 2020;125:1–9.
27. Ostrander EA, Dreger DL, Evans JM. Canine Cancer Genomics: Lessons for Canine and Human Health. *Annu Rev Anim Biosci*. 2019;7:449–72.
28. Savolainen P, Leitner T, Wilton AN, Matisoo-Smith E, Lundeberg J. A detailed picture of

- the origin of the Australian dingo, obtained from the study of mitochondrial DNA. *Proc Natl Acad Sci U S A*. 2004;101:12387–90.
29. Duleba A, Skonieczna K, Bogdanowicz W, Malyarchuk B, Grzybowski T. Complete mitochondrial genome database and standardized classification system for *Canis lupus familiaris*. *Forensic Sci Int Genet*. 2015;19:123–9.
30. Vonholdt BM, Pollinger JP, Lohmueller KE, Han E, Parker HG, Quignon P, et al. Genome-wide SNP and haplotype analyses reveal a rich history underlying dog domestication. *Nature*. 2010;464:898–902.
31. Hedan B, Cadiou E, Botherel N, Dufaure de Citres C, Letko A, Rimbault M, et al. Identification of a Missense Variant in MFSD12 Involved in Dilution of Phaeomelanin Leading to White or Cream Coat Color in Dogs. *Genes (Basel)*. 2019;10.
32. Berg P, Nielsen J, Sørensen MK. EVA: realized and predicted optimal genetic contributions. In: *Proceedings of the 8th World Congress on Genetics Applied to Livestock Production, Belo Horizonte, Minas Gerais, Brazil, 13-18 August, 2006*. 2006. p. 27-09.
33. Oliehoek PA, Bijma P, Van Der Meijden A. History and structure of the closed pedigreed population of Icelandic Sheepdogs. *Genet Sel Evol*. 2009;41:1–12.
34. Leonberger Health Questionnaire. Institute of Genetics, University of Bern, Canine Genetics Lab, University of Minnesota. https://www.genetics.unibe.ch/services/dog/european_leonberger_health_questionnaire/index_eng.html. Accessed 1 Jan 2020.
35. Chang CC, Chow CC, Tellier LCAM, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: Rising to the challenge of larger and richer datasets. *Gigascience*. 2015;4:1–16.
36. Lencz T, Lambert C, DeRosse P, Burdick KE, Morgan T V., Kane JM, et al. Runs of homozygosity reveal highly penetrant recessive loci in schizophrenia. *Proc Natl Acad Sci*. 2007;104:19942–7.
37. R Core Team. R: A language and environment for statistical computing. 2019.
38. Yin L. CMplot: Circle Manhattan Plot. 2019. <https://github.com/YinLiLin/R-CMplot>. Accessed 28 Feb 2020.
39. Lemon J. Plotrix: a package in the red light district of R. *R-News*. 2006;6:8–12.
40. Turner SD. qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. *bioRxiv*. 2014.
41. Online Mendelian Inheritance in Man, OMIM. McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University (Baltimore, MD). <https://omim.org/>. Accessed 28 Feb 2020.
42. Online Mendelian Inheritance in Animals, OMIA. Sydney School of Veterinary Science. <https://omia.org/>. Accessed 28 Feb 2020.
43. Choi Y, Chan AP. PROVEAN web server: A tool to predict the functional effect of amino acid substitutions and indels. *Bioinformatics*. 2015;31:2745–7.
44. Pejaver V, Urresti J, Lugo-Martinez J, Pagel KA, Lin GN, Nam H-J, et al. MutPred2: inferring the molecular and phenotypic impact of amino acid variants. *bioRxiv*. 2017;134981.
45. Pagel KA, Antaki D, Lian A, Mort M, Cooper DN, Sebat J, et al. Pathogenicity and functional impact of non-frameshifting insertion/deletion variation in the human genome. *PLOS Comput Biol*. 2019;15:e1007112.
46. Pagel KA, Pejaver V, Lin GN, Nam H-J, Mort M, Cooper DN, et al. When loss-of-function is loss of function: assessing mutational signatures and impact of loss-of-function genetic variants. *Bioinformatics*. 2017;33:i389–98.
47. Bendl J, Stourac J, Salanda O, Pavelka A, Wieben ED, Zendulka J, et al. PredictSNP: Robust and Accurate Consensus Classifier for Prediction of Disease-Related Mutations. *PLOS Comput Biol*. 2014;10:e1003440.
48. Pereira L, Van Asch B, Amorim A. Standardisation of nomenclature for dog mtDNA D-loop: A prerequisite for launching a *Canis familiaris* database. *Forensic Sci Int*. 2004;141:99–

108.

49. Thorvaldsson H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): High-performance genomics data visualization and exploration. *Brief Bioinform.* 2013;14:178–92.
50. Kerkel, K., Schupf, N., Hatta, K., Pang, D., Salas, M., Kratz, A., et al. Altered DNA methylation in leukocytes with trisomy 21. *PLoS genetics*, 2010;6:e1001212.
51. Zhang Z, Bai M, Barbosa GO, Chen A, Wei Y, Luo S, et al. Broadly conserved roles of TMEM131 family proteins in intracellular collagen assembly and secretory cargo trafficking. *Sci Adv.* 2020;6:eaay7667.
52. Leroy G, Phocas F, Hedan B, Verrier E, Rognon X. Inbreeding impact on litter size and survival in selected canine breeds. *Vet J.* 2015;203:74–8.
53. Lampi S, Donner J, Anderson H., Pohjoismäki J. Variation in breeding practices and geographic isolation drive subpopulation differentiation, contributing to the loss of genetic diversity within dog breed lineages. *Canine Med Genet* 2020;7:5.
54. Karlsson EK, Sigurdsson S, Ivansson E, Thomas R, Elvers I, Wright J, et al. Genome-wide analyses implicate 33 loci in heritable dog osteosarcoma, including regulatory variants near CDKN2A/B. *Genome Biol.* 2013;14:R132.
55. Sakthikumar S, Elvers I, Kim J, Arendt ML, Thomas R, Turner-Maier J, et al. SETD2 Is Recurrently Mutated in Whole-Exome Sequenced Canine Osteosarcoma. *Cancer Res.* 2018;78:3421–31.
56. Smallwood LJ, Barsanti JA. Hypoadrenocorticism in a family of Leonbergers. *J Am Anim Hosp Assoc.* 1995;31:301–5.
57. Segalini V, Hericher T, Grellet A, Rosenberg D, Garnier F, Fontbonne A. Thyroid function and infertility in the dog: a survey in five breeds. *Reprod Domest Anim.* 2009;44 Suppl 2:211–3.
58. Wiberg M, Niskanen JE, Hytönen M, Dillard K, Hagner K, Anttila M, et al. Ventricular arrhythmia and sudden cardiac death in young Leonbergers. *J Vet Cardiol Off J Eur Soc Vet Cardiol.* 2020;27:10–22.
59. Shelton GD, Podell M, Poncelet L, Schatzberg S, Patterson E, Powell HC, et al. Inherited polyneuropathy in Leonberger dogs: A mixed or intermediate form of Charcot-Marie-Tooth disease? *Muscle and Nerve.* 2003;27:471–7.
60. Leonberger Health Survey 2019. Leonberger Health Foundation International. https://www.leohealth.org/images/surveys-research/LHFI_HealthSurvey_PresentationMay2019.pdf. Accessed 1 Jan 2020.
61. Baranowska I, Jäderlund KH, Nennesmo I, Holmqvist E, Heidrich N, Larsson N-G, et al. Sensory ataxic neuropathy in golden retriever dogs is caused by a deletion in the mitochondrial tRNA^{Tyr} gene. *PLoS Genet.* 2009;5:e1000499.
62. Hultin Jäderlund K, Baranowska Körberg I, Nødtvedt a. Inherited Polyneuropathy in Leonberger Dogs. *J Vet Intern Med.* 2011;25:997–1002.

Figures

Figure 1 Pedigree analyses of Leonberger dogs.

A. Average pedigree-based inbreeding coefficient and kinship estimated per year from 1920 to 2016. B. Pie charts of the proportion of sires used in breeding clearly showing the popular sire syndrome. Popular sires are males that produced at least 33 puppies and account for 25% of all breeding males. These sires are responsible for 66% of all offspring. C. Boxplot of age at death per year of birth showing the life expectancy of Leonberger dogs over years 1989–2004. The exact number of records available per year is indicated in red above each year.

Figure 2 Graphical representation of the genomic analyses of Leonberger dogs.

A. MDS plot of 1203 dogs showing no obvious subpopulations regardless the geographical region of residence. B. Correlation of inbreeding coefficients estimated from the pedigree data (F_PED) and the ROH analysis (F_ROH). Note that the dogs with unknown pedigrees

(F_PED=0) on the left side mostly show the same level of inbreeding as the dogs with known registered records. C. Frequencies and counts of the different classes of ROH based on their length indicating both old and recent inbreeding. D. Number of detected ROHs per autosome (left y axis) and the proportion of chromosomes in a ROH (right y axis). E. Manhattan plot showing the proportion of dogs sharing a SNP within a ROH over the 38 autosomes. Most dogs share a region on chromosome 10 and 26.

Tables

Table 1 Rare protein-changing variants private in Leonberger dogs

Gene	OMIM/ OMIA number	Variant designation ¹			Alternative allele frequency	
		Genomic position	Coding DNA change	Protein change	Leonbergers	controls
<i>ATXN7</i>	607640	chr20:27234549	c.2285T>C	p.Val762Ala	0.1154	
<i>CEP55</i>	610000	chr28:7777489	c.1334A>C	p.Tyr446Ser	0.0256	
<i>CNTF</i>	118945	chr18:37758771	c.401_402insA	p.Asn137fs	0.0128	
<i>ELOVL4</i>	605512	chr12:40850011	c.424_433delGGAG CACAGC	p.Gly142fs	0.0128	
<i>FDX1L</i>	614585	chr20:50798661	c.124_125insGGCC GCCATCACCACG GCGGTGAGCACC GCCGCCAGCAGC ACCAGCCCGTCA GCGTTGAGCCG	p.Thr41_Ala4 2insGlyProPro SerProArgArg Ter	0.0128	
<i>GJA9</i> ²	611923/2 119-9615	chr15:3863519	c.1108_1109delGA	p.Glu370fs	0.0513	
<i>MCM3AP</i>	603294	chr31:39553337	c.1785_1793delCTC TGAAGG	p.Ser596_Gly5 98del	0.0128	
<i>NAPEPLD</i> ³	612334/1 788-9615	chr18:16987520	c.559G>C	p.Ala187Pro	0.1154	
<i>PLEKHG5</i>	611101	chr5:60325903	c.1585C>T	p.Gln529*	0.0128	
<i>SPTBN4</i>	606214/2 232-9823	chr1:113215064	c.1247_1248insGGT AGCCCATGCGGT	p.Ala416_Ala 417insValAla HisAlaVal	0.0128	
<i>SYNE1</i>	608441	chr1:42549994	c.17359C>T	p.Arg5788Trp	0.0256	

¹Additional details including the protein prediction effects are described in Additional file 10. All positions refer to the CanFam3.1 reference sequence assembly.

²Previously described polyneuropathy-associated variant [5].

³Previously described leukoencephalomyelopathy-associated variant [6].

Table 2 Protein-changing variants in polyneuropathy-associated candidate genes enriched in Leonberger dogs compared to 614 controls of various breeds

Gene	OMIM/ OMIA number	Variant designation ¹			Alternative allele frequency	
		Genomic position	Coding DNA change	Protein change	Leonbergers	controls
<i>CNTNAP1</i>	602346	chr9:20294320	c.3863G>C	p.Arg1288Pro	0.3974	0.0366
<i>CNTNAP1</i>	602346	chr9:20298261	c.2585G>A	p.Gly862Glu	0.1538	0.0067
<i>DHTKD1</i>	614984	chr2:24316951	c.1820C>T	p.Ala607Val	0.0897	0.0049
<i>DIAPH3</i>	614567	chr22:15880854	c.2000G>A	p.Cys667Tyr	0.1538	0.0689

<i>DST</i>	113810	chr12:23771235	c.18224C>T	p.Thr6075Met	0.0769	0.0008
<i>DST</i>	113810	chr12:23782332	c.17800C>T	p.Arg5934Trp	0.4359	0.0330
<i>DST</i>	113810	chr12:23846624	c.10661C>T	p.Ser3554Leu	0.0769	0.0369
<i>DYNC1H1</i>	600112	chr8:70064306	c.13757C>T	p.Pro4586Leu	0.1282	0.0051
<i>GARS</i>	600287	chr14:43322005	c.622G>A	p.Val208Ile	0.1154	0.0025
<i>GJA9</i>	611923/2 119-9615	chr15:3862761	c.344G>C	p.Arg115Thr	0.5513	0.0868
<i>JPH1</i>	605266	chr29:22710800	c.1531A>G	p.Ile511Val	0.1923	0.0299
<i>LOC477508</i>	602072	chr26:16348822	c.298G>A	p.Val100Met	0.1795	0.0066
<i>MME</i>	120520	chr23:49045461	c.1700A>T	p.Gln567Leu	0.0769	0.0157
<i>NEFH</i>	162230	chr26:22729485	c.1046G>A	p.Arg349His	0.2821	0.0116
<i>NEFH</i>	162230	chr26:22732974	c.1901_1924 delTGAAGG AGGAGGC CAAGTCCC CAG	p.Val640_Pro6 47del	0.3974	0.0849
<i>NEFH</i>	162230	chr26:22733415	c.2326C>G	p.Pro768Ala	0.0513	0.0059
<i>OTOF</i>	603681	chr17:20534866	c.3451G>A	p.Ala1151Thr	0.7051	0.0094
<i>PDXK</i>	179020	chr31:37707445	c.774G>C	p.Arg258Ser	0.1923	0.0898
<i>PRX</i>	605725	chr1:113290407	c.784G>A	p.Ala262Thr	0.1282	0.0257
<i>SBF2</i>	607697	chr21:33036988	c.4114C>T	p.Pro1372Ser	0.1667	0.0361
<i>SLC12A6</i>	604878	chr30:853540	c.2498C>T	p.Ala833Val	0.0513	0.0250
<i>WNK1</i>	605232	chr27:42911057	c.7448C>G	p.Thr2483Arg	0.0128	0.0058

¹Additional details including the protein prediction effects are described in Additional file 10.

All positions refer to the CanFam3.1 reference sequence assembly.

Table 3 Private variants in mitochondrial DNA of Leonberger dogs

Gene	Variant designation ¹	Proportion of variant ²	OMIM/OMIA number	Type
<i>tRNA-Phe</i>	m.49A>G	96%	590070	unknown
<i>tRNA-Val</i>	m.1069T>C	99%	590105	unknown
<i>ND2</i>	m.4764T>C; p.Ile284Thr	100%	516001	missense
<i>COX1</i>	m.5681A>G	66%, 89%	516030	synonymous
<i>ND4</i>	m.11211C>T	72%, 66%	516003	synonymous
<i>ND5</i>	m.13118T>C	62%	516005	synonymous
<i>ND5</i>	m.13544A>G	100%, 100%, 100%, 100%	516005	synonymous

¹All positions refer to the CanFam3.1 reference sequence assembly.

²The percentage of the variant represents heteroplasmy, 100% represents homoplasmy, for each dog that carries the variant calculated from the coverage of the affected mtDNA region.

Additional files

Additional file 1

Format: xlsx

Title: List of 1,203 studied dogs with available genomic information.

Description: The table includes supplemental descriptive details of the studied Leonberger dogs including accession numbers and known health history of the WGSed dogs.

Additional file 2

Format: xlsx

Title: List of 113 selected genes used for filtering variants in the 653 available WGS data.

Description: Putative candidate genes and their phenotype association in humans or animals based on OMIM [41] and OMIA [42] databases showing the number of variants detected in each of these genes in the Leonberger genomes.

Additional file 3

Format: tiff

Title: Population size of the Leonberger breed between the years 1920-2016.

Description: The blue, yellow, and green lines represent the number of breeding males, breeding females, and puppies they produced per year, respectively. The red dashed line represents the number of expected parent pairs for the given number of born puppies per year, assuming the average litter size of 6.5 puppies. Panel A shows the increasing population size from 1920 to 1978 and the apparent bottleneck around 1946. In the 1970s the number of born dogs started to increase rapidly while the number of used breeding males started to decrease. Panel B shows the continuously increasing population from 1979 to 2016. Note that the number of dams is more or less as expected but the number of sires constantly decreases to about half of the dams in recent years, illustrating the popular sire syndrome.

Additional file 4

Format: tiff

Title: Mean kinship of the estimated current population of 31,832 Leonberger dogs per country.

Description: The proportion of dogs belonging into the three groups, which indicate the increasing relatedness to the whole population, is shown as green ($MK < 0.31$), yellow ($MK = [0.31-0.32]$), and orange ($MK > 0.32$) per each country. The total number of dogs recorded in each country is shown in red on top of the corresponding columns.

Additional file 5

Format: xlsx

Title: Distribution of disease phenotypes reported in 2,726 Leonberger dogs.

Description: The table shows the number of dogs for which the owners reported particular disorders. Each more general group is divided in specific subgroups. Note that many dogs have multiple health issues and therefore the sum of the individual counts will be higher than the total of 1,334 dogs with at least one health condition reported.

Additional file 6

Format: tiff

Title: MDS plots of the 1,203 Leonberger dogs highlighted by different groups.

Description: Panel A shows the distribution of dogs per each country. Panel B shows the dogs divided into five groups by their year of birth. This is also reflected in panel C, where the dogs are coded by their calculated MK coefficient, which was not determined for the older dogs.

Additional file 7

Format: pdf

Title: Proportion of dogs with a SNP within a ROH on each of the 38 canine autosomes.

Description: Individual Manhattan plots showing the percentage of 1,203 SNP array genotyped dogs that share a SNP within a ROH over each canine autosome.

Additional file 8

Format: xlsx

Title: List of private whole-genome sequence variants in 39 Leonberger dogs.

Description: Breed-specific SNVs present at least heterozygous in all Leonberger dogs and absent from the 614 control genomes.

Additional file 9

Format: xlsx

Title: List of rare whole-genome sequence variants in 39 Leonberger dogs.

Description: Breed-specific SNVs present in at least one Leonberger dog and absent from the 614 control genomes.

Additional file 10

Format: xlsx

Title: List of the potentially pathogenic variants for neurological disorders detected in 39 Leonberger dogs including the predictions of variant effects on the proteins.

Description: The table includes detailed description and predictions of the biological consequences of the discovered variants on the corresponding proteins using different *in silico* prediction tools: PROVEAN [43], the MutPred suite (includes MutPred2 [44], MutPred-Indel [45], MutPred-LOF [46]), and PredictSNP [47].

Additional file 11

Format: xlsx

Title: List of whole-genome sequence variants present in at least one Leonberger dog in the region of 113 selected genes showing genotypes of all 653 available genomes.

Description: SNVs present in polyneuropathy- and Charcot-Marie-Tooth disease-associated candidate genes (listed in Additional file 2) and their alternative allele frequency in the Leonberger dogs and 128 other dog breeds and wolves.

Additional file 12

Format: xlsx

Title: Diversity of the mitochondrial genome of 39 Leonberger dogs.

Description: The table shows the low diversity of Leonberger mtDNA based on two nomenclature standards [29, 48]: most dogs belong to the A1b2a1a1 or A2 haplogroup while only a single dog belongs to the A1b1a1a or A17 haplogroup. Additionally, seven private SNVs detected in the mtDNA of the Leonbergers are also highlighted.