

Entwicklung eines *in silico* SNP-Arrays zur Genotypisierung von Hunden

Anton Ulbricht, Nils Schön, Röbbbe Wünschiers

Hochschule Mittweida

1. Hintergrund

Schon seit Jahrtausenden wird der Hund (*Canis lupus familiaris*) vom Menschen gezüchtet, um uns als Jagdgefährte, Hüte- und Wachhund oder Familienmitglied zur Seite zu stehen (Perri et al., 2021). Durch diese gezielte Selektion bezüglich eines bestimmten Verhaltens oder besonderen morphologischen Eigenschaften entstanden über die Zeit viele verschiedene Hunderassen. Dabei ist ihre genaue Herkunft nicht immer geklärt. Ein Beispiel dafür ist der Altdeutsche Schafpudel, welcher vermutlich aus der ostdeutschen Tiefebene stammt und dort vor allem als Hütehund eingesetzt wurde. Die Population der Altdeutschen Schafpudel in Deutschland verkleinerte sich jedoch ab Ende der 1980er Jahre durch Umstrukturierungen im Schäfererwesen erheblich. Heute wird er vermehrt als Familien- und Therapie- sowie Rettungshund gezüchtet. Um den Altdeutschen Schafpudel phylogenetisch einordnen und eventuelle Rückschlüsse auf seine Herkunft ziehen zu können, erfolgte die *in silico* Genotypisierung anhand von Sequenzierungsdaten eines Schafpudel-Rüden. Für die phylogenetische Analyse wurden 729 weitere Hunde herangezogen, so dass die Analyse insgesamt 130 verschiedenen Hunderassen umfasste (Dutrow et al., 2022; Parker et al., 2017).

2. Methoden

Basierend auf den Sondensequenzen des CanineHD BeadChip von Illumina (Illumina, 2023) wurde im Rahmen dieses Projektes ein *in silico* SNP-Array zur Genotypisierung von Hunden ausgehend von Illumina *short reads* entwickelt. Diese Reads wurden mittels der Solex Sequenzierertechnologie von Illumina erzeugt. Mithilfe des CanineHD BeadChip wurden bereits zahlreiche Hunde unterschiedlichster Rassen genotypisiert, daher ist der resultierende Genotyp direkt mit einer großen Anzahl an verfügbaren Daten vergleichbar.

Zu Beginn wurde aus den 50 bzw. 51 bp langen Sondensequenzen eine Datenbank als Grundlage für ein Alignment erzeugt. Dabei wurde jeweils die Position, an der sich der SNP befinden sollte, mit einem N markiert. Anschließend wurde ein Alignment der ca. 150 bp langen Short-Reads mit allen Sondensequenzen des SNP-Arrays durchgeführt. Alignments mit einer Mindestlänge von 50 bp und einer Identität von 100% wurden für die weitere Analyse verwendet. Des Weiteren wurden alle SNPs aus den Alignments isoliert und als Genotyp zusammengefasst. Anschließend wurden für jede Variante die vorliegenden Allele gezählt und der Genotyp dieser Variante bestimmt. Varianten, bei denen nur ein einziges Allel gefunden werden konnte, wurde als homozygot in diesem Allel angegeben. Wurden zwei unterschiedliche Allele gefunden, wurde die Variante als heterozygot bestimmt und wurden keine, drei oder vier verschiedene Allele für eine Variante gefunden, wurde der Genotyp auf „missing“ bzw. „0 0“ gesetzt, da das Ergebnis nicht eindeutig war. Zwei Optimierungsschritte wurden durchgeführt, um die Anzahl spezifischer Alignments zu erhöhen und somit die größtmögliche Menge an Informationen aus den Reads zu erhalten. Einerseits wurden dabei die *reverse complements* der

Sondensequenzen mit in die Datenbank einbezogen, wodurch die Anzahl der Alignments verdoppelt werden konnte. Des Weiteren wurde die Länge der Sondensequenz auf 20 bp reduziert, um weitere Alignments zu ermöglichen und trotzdem eine ausreichend hohe Spezifität zu gewährleisten. Der resultierende Genotyp des Altdeutschen Schafpudels wurde mit den 729 anderen Genotypen, die mit Hilfe des CanineHD Bead-Chips bestimmt wurden, zu einer gemeinsamen .ped-Datei zusammengefügt (Dutrow et al., 2022; Parker et al., 2017). Danach wurde eine Identity-By-State (IBS) Distanzmatrix mit PLINK 1.9 berechnet (Chang et al., 2015). Diese Distanzmatrix wurde wiederum dafür verwendet, um mit Hilfe von PHYLIP 3.697 (Felsenstein, 1989) ein Kladogramm zu konstruieren, dessen Signifikanz abschließend mittels Bootstrap-Analyse im Umfang von 100 Samples validiert wurde.

3. Ergebnisse

Mit der ursprünglichen Methode des *in silico* SNP-Arrays konnten Informationen über 157.422 Varianten von insgesamt 173.662 möglichen Varianten, die sich auf dem originalen BeadChip befinden, gesammelt werden. Die beiden Optimierungsschritte ermöglichten es, die Anzahl von Varianten, für die nach dem Alignment Informationen vorliegen, auf 173.271 zu steigern. In das resultierende Kladogramm konnten alle 730 Individuen eingeordnet werden, wobei Hunde derselben Rasse bis auf wenige Ausnahmen immer innerhalb einer Klade eingeordnet wurden. Der Bootstrap-Wert eines Knotenpunktes gibt an, in wie vielen Fällen die Anordnung rechts des Knotenpunktes aufgetreten ist. Ein Wert von 100 bedeutet, dass die jeweilige Konstellation in allen Bootstrap Samples exakt so vorgefunden werden konnte. Ab einem Wert von 70 kann von einem signifikanten Ergebnis ausgegangen werden (Hillis & Bull, 1993).

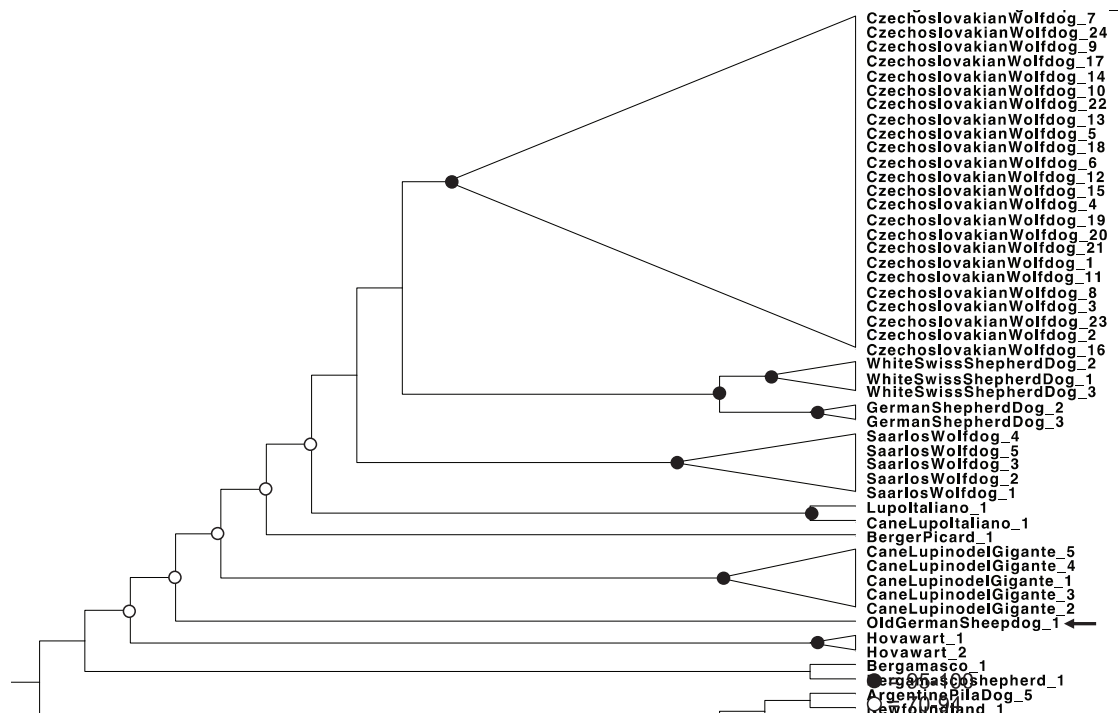


Abbildung 1: Ausschnitt des Kladogramms. Der Altdeutsche Schafpudel (engl.: Old German Sheepdog) ist mit einem Pfeil markiert. Weiß gefüllte Kreise markieren Knotenpunkte mit Bootstrap-Werten zwischen 70 und 94, schwarz gefüllte Kreise Knotenpunkte mit Werten größer gleich 95. Knotenpunkte ohne farbigen Kreis besitzen einen Bootstrap-Wert unter 70.

Im gesamten Kladogramm sind viele kleinere Gruppen mit einem hohen Bootstrap-Wert zu finden. In Abbildung 1 ist ein Ausschnitt des aus der Bootstrap-Analyse resultierenden Kladogramms dargestellt. Zu erkennen ist, dass alle Hunde derselben Rasse jeweils mit einem Bootstrap-Wert von über 95 sehr zuverlässig in eine gemeinsame, monophyletische Klade eingeordnet werden konnten. Der Altdeutsche Schafpudel ist in einer Klade für sich eingeordnet, mit dem nächsten Verwandten Cane Lupino Del Gigante, einem italienischen Hütehund. Die hier gezeigten nah verwandten Hunderassen des Altdeutschen Schafpudels stammen aus Italien (Lupo Italiano, Lupino del Gigante), der Schweiz (White Swiss Shepherd Dog), Frankreich (Berger Picard), den Niederlanden (Saarloos Wolfdog), Deutschland (German Shepherd Dog, Hovawart) und dem tschechoslowakischen Raum (Czechoslovakian Wolfdog).

4. Diskussion

Die hier gezeigten Ergebnisse deuten darauf hin, dass die Genotypisierung mittels *in silico* SNP-Array für den Schafpudel funktionierte und der Schafpudel scheinbar sinnvoll eingeordnet wurde. Da der Bootstrap-Wert über 70 liegt und somit signifikant ist (Hillis & Bull, 1993), lässt sich schlussfolgern, dass der Altdeutsche Schafpudel in dem hier betrachteten Datensatz am nächsten zu Lupino del Gigante verwandt ist. Der Lupino del Gigante stammt aus Italien, jedoch ist seine Herkunft nicht eindeutig geklärt. Der Altdeutsche Schafpudel weist ebenso eine hohe genetische Ähnlichkeit mit den anderen Rassen auf, die sich rechts des Knotenpunktes befinden. Sowohl Saarloos Wolfdog, Czechoslovakian Wolfdog als auch vermutlich Lupo Italiano sind Rassen, die durch die Zucht eines German Shepherd Dog mit einem Wolf entstanden sind (Fédération Cynologique Internationale, 2023; Petzold, 2023). Diese drei Rassen sind mit einem Bootstrap-Wert von über 70 in einer gemeinsamen Klade eingeordnet, was die allgemeine Qualität der verwendeten Daten und die korrekte Konstruktion des Kladogramms bestätigt. Dies kann auch anhand weiterer Beispiele im Kladogramm überprüft werden. Der German Shepherd Dog zeigt eine hohe Präsenz in der Klade, ebenfalls ist er auch ein Vorfahre des Hovawart. Somit ist davon auszugehen, dass er auch als möglicher Vorfahre des Altdeutschen Schafpudels in Frage kommen könnte. Dass die nächsten Verwandten im phylogenetischen Baum aus Italien und Deutschland stammen, deutet daraufhin, dass sich die Herkunft des Altdeutschen Schafpudels in eine dieser Regionen einordnen lässt. Abschließend beruht die hier dargestellte Analyse jedoch auf einem einzigen Individuum und einem begrenzten Datensatz, weshalb noch keine allgemeingültige Aussage über die Herkunft der Rasse getroffen werden kann.

5. Ausblick

Die hier gezeigten Ergebnisse könnten durch Einbeziehen weiterer bereits verfügbarer Genotypen anderer Hunderassen ergänzt werden, um ein detailreicheres Kladogramm zu erhalten. Des Weiteren wäre es aufschlussreich, weitere Schafpudel zu genotypisieren und deren Genotyp in die phylogenetische Analyse einzubeziehen. Um die Funktion des SNP-Arrays vollständig zu validieren, wäre es außerdem notwendig, einen Hund auf mit dem *in silico* SNP-Array zu genotypisieren, für den bereits andere Individuen derselben Rasse oder er selbst mit Hilfe des Canine HD BeadChip genotypisiert wurden. Würde dieser Hund dann gemeinsam mit den anderen in eine monophyletische Klade eingeordnet werden, oder würden sich die beiden erzeugten Genotypen desselben Hundes gleichen, würde das die Funktionsfähigkeit des *in silico* SNP-Arrays bestätigen. Weiterführend könnte der durch die hier angewandte Methode erzeugte Genotyp für

weitere Analysen neben der phylogenetischen Analyse verwendet werden. Zeitgleich wurden die zugrundeliegenden *Sequencing Reads* auf das CanFam 3.0 Referenzgenom gemappt und die SNPs gecallt. Die Ergebnisse dieses Verfahrens könnten ebenfalls mit denen des *in silico* SNP-Arrays verglichen werden, um die Qualität dieser Methode an aktuellen bioinformatischen Tools zu messen.

Quellen

Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., & Lee, J. J. (2015). Second-generation PLINK: Rising to the challenge of larger and richer datasets. *Giga-Science*, 4(1), 7. <https://doi.org/10.1186/s13742-015-0047-8>

Dutrow, E. V., Serpell, J. A., & Ostrander, E. A. (2022). Domestic dog lineages reveal genetic drivers of behavioral diversification. *Cell*, 185(25), 4737-4755.e18. <https://doi.org/10.1016/j.cell.2022.11.003>

Fédération Cynologique Internationale. (2023, Januar 26). *Fci.be*. <https://www.fci.be/en/>

Felsenstein, J. (1989). PHYLIP: Phylogeny Inference Package. Version 3.2. *Cladistics*, 5, 164–166. <https://doi.org/10.1086/416571>

Hillis, D. M., & Bull, J. J. (1993). An Empirical Test of Bootstrapping as a Method for Assessing Confidence in Phylogenetic Analysis. *Systematic Biology*, 42(2), 182. <https://doi.org/10.2307/2992540>

Illumina. (2023, Januar 26). *CanineHD BeadChip probe sequences*. <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GPL20953>

Parker, H. G., Dreger, D. L., Rimbault, M., Davis, B. W., Mullen, A. B., Carpintero-Ramirez, G., & Ostrander, E. A. (2017). Genomic Analyses Reveal the Influence of Geographic Origin, Migration, and Hybridization on Modern Dog Breed Development. *Cell Reports*, 19(4), 697–708. <https://doi.org/10.1016/j.celrep.2017.03.079>

Perri, A. R., Feuerborn, T. R., Frantz, L. A. F., Larson, G., Malhi, R. S., Meltzer, D. J., & Witt, K. E. (2021). Dog domestication and the dual dispersal of people and dogs into the Americas. *Proceedings of the National Academy of Sciences*, 118(6), e2010083118. <https://doi.org/10.1073/pnas.2010083118>

Petzold, F. (2023). *Lupo Italiano – Wolfshybrid & Wolfshund Hunderassen*. <https://der-wolfshund.de/wolfshund-hunderassen/lupo-italiano>