

Genetic Analysis of Human Evolution:
The Case of Oceania

Inaugural-Dissertation

zur
Erlangung des Doktorgrades
der Mathematisch-Naturwissenschaftlichen Fakultät
der Universität zu Köln

vorgelegt von
Andreas Wollstein
aus Berlin

Köln, 2011

Berichtersteller: Prof. Dr. Peter Nürnberg
Prof. Dr. Thomas Wiehe

Tag der letzten mündlichen Prüfung: 6. Dezember 2011

Abstract

Studying genetic variation of human populations is interesting from at least two perspectives. First, it allows us to do inferences about the history of populations, which includes changes in population size and migration events. Second, it is known that adaptive evolution leaves certain footprints on the genome, which can be used to identify such regions. The knowledge about differences in the evolution of functionally relevant genetic regions might reveal new insights into population specific predisposition to certain diseases.

Here the human genetic history with the example of Oceania is investigated. The region of Oceania is distinguished into Near Oceania (including New Guinea, the Solomon Islands, up to Santa Cruz Islands) and Remote Oceania (including New Caledonia, Vanuatu, Fiji, and Polynesia), which is of interest, because it represents the most recent human occupation of a large previously uninhabited geographic region of the world. The settlement history of Oceania is assumed to have happened via two major waves. Archaeological evidence suggests that the first settlement has reached Near Oceania about 40 thousand years ago. The Austronesian expansion originating from East Asia about 6000 years ago has fueled a second migration wave that went through Near Oceania into Remote Oceania. During the traverse of Near Oceania, a considerable amount of admixture with the aboriginal people was achieved as genetic evidence using uniparentally inherited markers suggest. The Austronesian culture that early-comprised skills of seafarers and navigation allowed them to rapidly traverse Near Oceania with an unprecedented speed of a few thousand years. It has been assumed that during long periods of famine, genes have been selected that enabled the metabolism of food more efficiently, which might render a detrimental effect today due to the abundance of food that is easily available. This mechanism known as the thrifty gene hypothesis was often utilized to explain the high obesity ratio in Polynesian populations, particularly those living in western conditions, compared to other populations.

Details about the migration history in Oceania are still a matter of debate, which has been conducted mostly by archaeological, linguistically, and fewer genetic arguments. Genetic studies are currently underrepresented in Oceania and are mostly based on uniparental markers such as mitochondrial DNA or non-recombining Y chromosomal DNA. To study the demographic and selective forces that have shaped the genetic variation in Oceania, Borneans, Polynesians, New Guineans, and Fijians have been genotyped in about one million single nucleotide polymorphisms. This data set was extended by publicly available data from populations representing Africa, Europe and Asia. It has been previously criticized that model free statistical methods such as principal component analysis, used to reveal population sub-structuring, often lead to ambiguous estimates

about the underlying evolutionary forces that actually shaped the observed genetic variation. Therefore a model-based approach was chosen, namely Approximate Bayesian Computation to disentangle various demographic models and estimate its parameters. A new approach has been developed that deals with the SNP ascertainment bias and enables unbiased parameter estimates.

The results of the demographic investigation show that the genetic variation within Oceania compared to other populations overall, confirms the out of Africa theory according to which one major population migration has been diverged between 55-100 thousand years ago. Model comparisons revealed New Guinea to have split from European and Asian ancestors as a separate migration into Oceania about 27 thousand years ago. The successive Austronesian expansion has been estimated to result in a mixed Asian (87%) and New Guinean (13%) ancestry about three thousand years ago, which is confirming previous findings. The second study aims to analyze regions on the genome that are shaped by positive selection. To better distinguish between effects caused by population history and adaptation, existing approaches have been refined and extended in two ways. First, simulations of the test statistic under the most likely derived demographic model were used to employ more stringent classification borders. Second, ascertained regions under selection that were shared with any of the ancestral populations were excluded. As a result, Polynesians exhibit the highest number of selective sweep regions compared to other populations from Oceania. Furthermore, Polynesians exhibit an enrichment of genes located in selective regions that have been associated with obesity related phenotypes. The results from that study provide new evidence for the thrifty gene hypothesis in Polynesia.

Overall this genome-wide data investigation provides new insights into the demographic and adaptation history of Oceania that allow the answering of a number of previously open questions and at the same time opens new directions for future research.

Zusammenfassung

Die Analyse genetischer Variation des Menschen ist aus zweierlei Hinsicht interessant: Einerseits kann das durch neutrale Evolution geprägte Muster der genetischen Variabilität Rückschlüsse auf die Besiedlungsgeschichte liefern, die im Wesentlichen durch demographische Parameter wie die Populationsgröße, Migration und Isolation bestimmt ist. Andererseits hinterlassen rezente Selektionsereignisse lokale Spuren auf dem Genom, welche in populationsgenetischen Studien identifiziert werden können. Die Kenntnis über die differenzierte Entwicklung funktionsrelevanter genetischer Regionen, die beispielsweise durch positive Selektion verursacht wurde, liefert mögliche Hinweise über medizinisch relevante Veranlagungen bestimmter Phänotypen. In der vorliegenden Arbeit werden adaptive und neutrale evolutionäre Prozesse am Beispiel von Ozeanien untersucht.

Geographisch wird Ozeanien in Nah- und Fernozeanien unterteilt. Nahozeanien umfasst im Wesentlichen Neuguinea bis zu den Solomon- und Santa-Cruz-Inseln. Fernozeanien umfasst die Inseln östlich der Santa-Cruz-Inseln: Neukaledonien, Vanatu, Fiji und Polynesien, das noch bis vor ungefähr 3000 Jahren von Menschen völlig unbesiedelt war. Es wird vermutet, dass Ozeanien in zwei Phasen besiedelt wurde. Archäologische Hinweise deuten darauf hin, dass die Besiedlung von Nahozeanien vor ungefähr 40000 Jahren stattgefunden hat. Die zweite Besiedlungsphase erfolgte danach durch die austronesische Expansion, die ihren Ursprung vermutlich vor ungefähr 6000 Jahren in Asien, Neuguinea nahm, die polynesischen Inseln durchquerte und vor ungefähr 3000 Jahren Fernozeanien erreichte. Analysen von uniparentalen Markern, wie zum Beispiel, mitochondriale DNA und Y-chromosomale DNA, zeigen eine beträchtliche Vermischung der Austronesen mit den Ureinwohnern Neuguineas.

Die in der austronesischen Kultur früh entwickelten Techniken der Seefahrt und Navigation ermöglichten desweiteren eine relativ schnelle Durchquerung des Archipels im Zeitraum von wenigen tausend Jahren. Es wird angenommen, dass sich durch häufige Hungerperioden solche Gene im Darwinschen Sinne durchsetzten, die eine effizientere Verwertung der gegebenen Nahrungsressourcen ermöglichten. Dagegen entwickelt sich dieser damalige "Überlebensvorteil" in vielen Regionen der westlichen Welt in heutigen Zeiten des Überangebots an Nahrung zum Nachteil. Dieser Mechanismus, der als „Thrifty Gene“-Hypothese bekannt ist, wurde häufig verwendet, um die im Vergleich zu anderen Populationen überdurchschnittlich hohe Adipositasrate in der polynesischen Bevölkerung zu erklären.

Viele Details der Besiedlungsgeschichte in Ozeanien sind noch im Unklaren und werden meist mit archäologischen und linguistischen Argumenten diskutiert. Genetische

Untersuchungen sind für diese Region derzeit noch wenig verfügbar und basieren oft nur auf uniparentale Markern wie mitochondriale- oder Y-chromosomale DNA. Um die Auswirkungen evolutionärer Kräfte auf die genetische Variabilität in Ozeanien zu untersuchen, wurden Bornesen, Polynesier, Neuguinesen und Fidschi in etwa einer Million Single-Nucleotid-Polymorphismen (SNP) typisiert. Zum Vergleich wurden öffentlich verfügbare Genotypen von Proben aus Afrika, Europa und Asien hinzugefügt.

Es wurde kritisiert, dass modellfreie Methoden, wie zum Beispiel die Hauptkomponentenanalyse, die zur Detektion von Substrukturen herangezogen werden, oft keine eindeutigen Ergebnisse zur zugrundeliegenden demographischen Historie liefern. Für die Untersuchung der neutralen Evolution wurde daher ein modellbasierter Ansatz gewählt. Mit Hilfe des „Approximate Bayesian Computation“-Verfahrens wurden verschiedenen demographische Modelle gegeneinander getestet und deren Parameter abgeschätzt. Ein neuer, aussichtsreicher Ansatz wurde entwickelt, der es erlaubt, solche demographischen Parameter von technischen Messeffekten („SNP-Ascertainment bias“) zu befreien.

Ergebnisse der Analysen neutraler Evolution bestätigen im Wesentlichen die Hypothese eines gemeinsamen Ursprungs der Menschheit in Afrika, der auf vor etwa 55000 bis 100000 Jahren datiert wird. Aus Modellvergleichen geht desweiteren hervor, dass sich die Vorfahren Neuguineas höchstwahrscheinlich aus den europäischen und asiatischen Vorfahren vor ungefähr 27000 Jahre abgespalten haben. Die Parameter der austronesischen Expansion ergeben eine Vermischung von Bornesen (87%) und Neuguinesen (13%) vor ungefähr 3000 Jahren, was frühere Ergebnisse aus uniparentalen Markern bestätigt.

Eine zweite Studie zielt auf die Detektion genomischer Regionen, die durch positive Selektionsereignisse geprägt wurden. Um besser zwischen den neutralen und adaptiven Signaturen zu unterscheiden, wurden vorhandene Ansätze verfeinert und erweitert. Zum einen wurden Simulationen verwendet, die eine bessere Abgrenzung von neutralen und adaptiven Signaturen ermöglicht. Zum anderen werden jene Selektionssignale, die auch in den ancestralen Population detektiert wurden, ausgeschlossen. Im Vergleich zu anderen Populationen aus Ozeanien weisen die Polynesier die höchste Anzahl von Regionen auf, für die eine Selektionssignatur nachgewiesen werden konnte. Darüber hinaus zeigen selektierte Gene in Polynesien eine vermehrte Assoziation mit Adipositas- und Adipositas-verwandten Phänotypen auf. Die Ergebnisse aus dieser Studie liefern somit neue Hinweise für die „Thrifty Gene“-Hypothese in Polynesien.

Insgesamt liefert die Arbeit aus der Untersuchung genomweiter Daten neue Einblicke in die demographische Geschichte Ozeaniens, die eine Beantwortung einer Reihe bisher offener Fragen erlaubt und gleichzeitig neue Wege für zukünftige Forschungsarbeit bietet.