

Abstract

Since the seminal work of Darwin and the modern evolutionary synthesis, population genetics has represented a powerful mathematical framework to understand evolution in natural populations. Over the past two decades, advances in sequencing technology have greatly expanded population genetics studies across a wide range of biological systems. These studies revealed general patterns of genomic diversity, the effects of evolutionary dynamics on genomic variation, and genes and variants associated with trait variation and disease. In this study, I applied population genetics approaches to study the evolutionary history of *Arabis alpina*, a perennial herb related to the annual model plant *Arabidopsis thaliana*. *A. alpina* has a broad Arctic-alpine distribution, and has been shown to be locally adapted with reciprocal transplant experiments. This species also displays high phenotypic variation in traits related to flowering, to the mating system, and to the harsh Arctic-alpine environment. Extensive research in phylogeography and in developmental biology, along with an increasing number of genetic and genomic resources, make this perennial, Arctic-alpine plant an excellent model to study population genetics and evolution. In this work, I present a comprehensive and novel genomic diversity panel of individual *A. alpina* plants (accessions) collected across Europe. This dataset provides a unique opportunity to investigate both the demographic and adaptive histories of *A. alpina* populations. Importantly, large-scale genomic datasets like the one presented here can be biased by complex genomic regions, such as repetitive elements and structural variants. In particular, we found that genomic regions present in multiple copies (multicopy genomic regions) can result in major errors by aligning to each other (collapse), and introduce biases in genomic data and in population genomic inference. To address this, we developed a novel method to detect collapsed multicopy regions, which can be applied to *A. alpina*, and to other model or non-model species. This method revealed that a significant proportion of genetic variation in *A. alpina* is associated with multicopy regions, and in particular with multicopy transposable elements, mostly long terminal repeat (LTR) retrotransposons. After filtering multicopy regions and other complex genomic regions, we used population genetics inference to reconstruct the demographic history of European *A. alpina* populations. Our analyses revealed the ancient history of *A. alpina* in Europe, with population splits dating back as far as 400,000 years, alongside fluctuations in effective population size that correspond to the last glacial episode. These events have contributed to the high population structure observed between Central European, Iberian, and Scandinavian populations. Additionally, we reconstructed the adaptive history of populations of *A. alpina* from the Cantabrian Mountains (Northern Spain), motivated by their divergent flowering behavior. We identified multiple candidate genes involved in drought response and in perennial flowering, which likely contributed to the adaptation of Cantabrian populations to seasonal drought. Finally, we show that adaptation in these populations is likely tied to climatic changes following the last glacial maximum.

Zusammenfassung

Seit der bahnbrechenden Arbeit von Darwin und der modernen synthetischen Evolutionstheorie stellt die Populationsgenetik ein mächtiges mathematisches Rahmenwerk zur Erklärung der Evolution in natürlichen Populationen dar. In den letzten zwei Jahrzehnten haben Fortschritte in der Sequenzierungstechnologie die populationsgenetische Forschung auf eine Vielzahl biologischer Systeme erheblich ausgeweitet. Diese Studien haben allgemeine Muster genomischer Diversität, die Auswirkungen evolutionsdynamischer Prozesse auf genomische Variation sowie Gene und Varianten, die mit Merkmalsvariation und Krankheiten assoziiert sind, aufgezeigt. In dieser Studie habe ich populationsgenetische Ansätze angewandt, um die Evolutionsgeschichte von *Arabis alpina* zu untersuchen, einer krautigen, mehrjährigen Pflanze, die mit dem einjährigen Modellorganismus *Arabidopsis thaliana* verwandt ist. *A. alpina* hat eine weite arktisch-alpine Verbreitung und zeigt laut reziproken Transplantationsexperimenten lokale Anpassungen. Die Art weist zudem eine hohe phänotypische Variation in Merkmalen auf, die mit der Blüte, dem Fortpflanzungssystem und der Anpassung an die harsche arktisch-alpine Umwelt zusammenhängen. Umfangreiche Forschung in der Phylogeographie und Entwicklungsbiologie sowie eine zunehmende Zahl genetischer und genomischer Ressourcen machen diese mehrjährige, arktisch-alpine Pflanze zu einem ausgezeichneten Modell für populationsgenetische und evolutionäre Studien. In dieser Arbeit präsentiere ich ein umfassendes und neuartiges Panel genomischer Diversität von einzelnen *A. alpina*-Pflanzen (Accessions), die in ganz Europa gesammelt wurden. Dieses Datenset bietet eine einzigartige Gelegenheit, sowohl die demografische als auch die adaptive Geschichte europäischer *A. alpina*-Populationen zu untersuchen. Wichtig ist, dass groß angelegte genomische Datensätze wie dieser durch komplexe genomische Regionen, etwa repetitive Elemente und strukturelle Varianten, verzerrt werden können. Insbesondere haben wir festgestellt, dass genomische Regionen, die in mehreren Kopien vorliegen (Multikopien-Regionen), durch gegenseitige Fehlzuordnung beim Alignieren (Collapse) zu erheblichen Fehlern führen können und dadurch Verzerrungen in den Daten und in populationsgenetischen Inferenzmethoden verursachen. Um dem entgegenzuwirken, haben wir eine neue Methode entwickelt, um kollabierte Multikopien-Regionen zu identifizieren, die sowohl auf *A. alpina* als auch auf andere Modell- und Nichtmodellorganismen anwendbar ist. Diese Methode zeigte, dass ein erheblicher Anteil der genetischen Variation in *A. alpina* mit Multikopien-Regionen, insbesondere mit transponierbaren Elementen, meist sogenannten LTR-Retrotransposons (Long Terminal Repeats), assoziiert ist. Nach dem Filtern von Multikopien- und anderen komplexen genomischen Regionen haben wir populationsgenetische Inferenzmethoden verwendet, um die demografische Geschichte europäischer *A. alpina*-Populationen zu rekonstruieren. Unsere Analysen zeigten eine sehr alte Präsenz von *A. alpina* in Europa, mit Populationsaufspaltungen vor bis zu 400.000 Jahren sowie Schwankungen in der effektiven Populationsgröße, die mit der letzten Eiszeit korrespondieren. Diese Ereignisse haben zur ausgeprägten Populationsstruktur zwischen zentral-, iberisch- und skandinavisch-europäischen Populationen beigetragen. Darüber hinaus haben wir die adaptive Geschichte der *A. alpina*-Populationen aus dem Kantabrischen Gebirge (Nordspanien) rekonstruiert, angeregt durch deren divergierendes Blühverhalten. Wir identifizierten mehrere Kandidatengene, die an der Trockenheitsantwort und an der mehrjährigen Blüte beteiligt sind und vermutlich zur Anpassung der kantabrischen Populationen an saisonale Trockenheit beigetragen haben. Abschließend zeigen wir, dass diese Anpassung wahrscheinlich im Zusammenhang mit klimatischen Veränderungen nach dem letzten glazialen Maximum steht.