

Reconstructing the population and adaptive history of *Arabidopsis thaliana*

Inaugural-Dissertation

zur

Erlangung des Doktorgrades

der Mathematisch-Naturwissenschaftlichen Fakultät

der Universität zu Köln

vorgelegt von

Sophie-Asako Xerri

aus Boulogne-Billancourt, Frankreich

Köln, 2025

Abstract

Arabidopsis thaliana is the major model species in plant molecular biology. As such, describing and understanding its natural variation can provide insights into the ecological and evolutionary drivers of molecular variation. Currently, the 1001 Genomes Project represents a valuable resource with data on over 1000 worldwide accessions for such studies. This project however suffers from sampling bias: most accessions come from the northern hemisphere –mainly from Europe. Previously, our research group included individuals from the African continent and showed that both the variation there and the inferred history are distinct from Eurasia's. In the present project, we further expanded our sampling with the aim of better understanding the global species history of *A. thaliana*. We sequenced nearly 800 newly collected accessions from East Africa as well as other under-represented regions such as the Levant. Our novel diversity panel has a total of over 2800 accessions covering many environments previously not considered.

Analyses of population structure revealed that East Africans indeed represent a distinct lineage from Eurasians. I also identified additional anciently diverged “relict” populations. Within the Eurasian continent, novel samples from Asia further clarified the local geographic structure. I then sought to better characterise many of our new populations by investigating local structure, population genetic summary statistics, as well as histories of selection and mutations. Lastly, I attempted to infer the global demographic history by employing coalescent-based methods to estimate split times at different scales, ranging from between major geographic regions to within countries. I suggest a new model for the species history of *A. thaliana*, with a single colonisation and gradual expansion in East Africa after characterising candidates for the ancestral population.

Our study thus sheds new lights onto the evolutionary history of *A. thaliana* on a more global scale, that could form a basis for further studies of local adaptation.

Zusammenfassung

Arabidopsis thaliana ist die wichtigste Modellart in der Pflanzenmolekularbiologie. Daher kann die Beschreibung und das Verständnis ihrer natürlichen Variation Erkenntnisse über die ökologischen und evolutionären Triebkräfte der molekularen Variation liefern. Derzeit stellt das *1001 Genomes Project* eine wertvolle Ressource mit Daten zu über 1000 weltweiten Akzessionen für solche Studien dar. Dieses Projekt leidet jedoch unter Stichprobenverzerrungen: die meisten Akzessionen stammen aus der nördlichen Hemisphäre – hauptsächlich aus Europa. Zuvor hatte unsere Forschungsgruppe Individuen vom afrikanischen Kontinent einbezogen und gezeigt, dass sich sowohl die dortige Variation als auch die daraus abgeleitete Geschichte von der Eurasiens unterscheidet. Im vorliegenden Projekt haben wir unsere Probenahme weiter ausgedehnt mit dem Ziel, die globale Artengeschichte von *A. thaliana* besser zu verstehen. Wir haben fast 800 neu gesammelte Akzessionen aus Ostafrika sowie aus anderen unterrepräsentierten Regionen, wie der Levante, sequenziert. Unser neuartiges Diversity-Panel umfasst insgesamt über 2800 Akzessionen, die viele bisher nicht berücksichtigte Umgebungen abdecken.

Analysen der Populationsstruktur ergaben, dass die ostafrikanischen Akzessionen tatsächlich eine von den eurasischen getrennte Abstammungslinie darstellt. Außerdem habe ich zusätzliche, vor langer Zeit divergierende, „Relict“-Populationen identifiziert. Innerhalb des eurasischen Kontinents brachten neue Proben aus Asien weitere Klarheit über die lokale geografische Struktur. Anschließend habe ich versucht, viele unserer neuen Populationen besser zu charakterisieren, indem ich die lokale Struktur, populationsgenetische Statistiken sowie die Geschichte von Selektion und Mutationen untersucht habe. Schließlich habe ich versucht, die globale demografische Geschichte abzuleiten, indem ich auf Koaleszenz-Theorie basierende Methoden einsetzte, um Trennungszeiten auf verschiedenen Skalen zu schätzen, von großen geografischen Regionen bis innerhalb von einzelnen Ländern. Ich schlage ein neues Modell für die Artengeschichte von *A. thaliana* vor, das eine einmalige Besiedelung und allmähliche Ausbreitung in Ostafrika vorsieht, nachdem ich Kandidaten für die angestammte Population charakterisiert habe.

Unsere Studie wirft somit ein neues Licht auf die Evolutionsgeschichte von *A. thaliana* auf einer globaleren Ebene, die eine Grundlage für weitere Studien zur lokalen Anpassung bilden könnte.