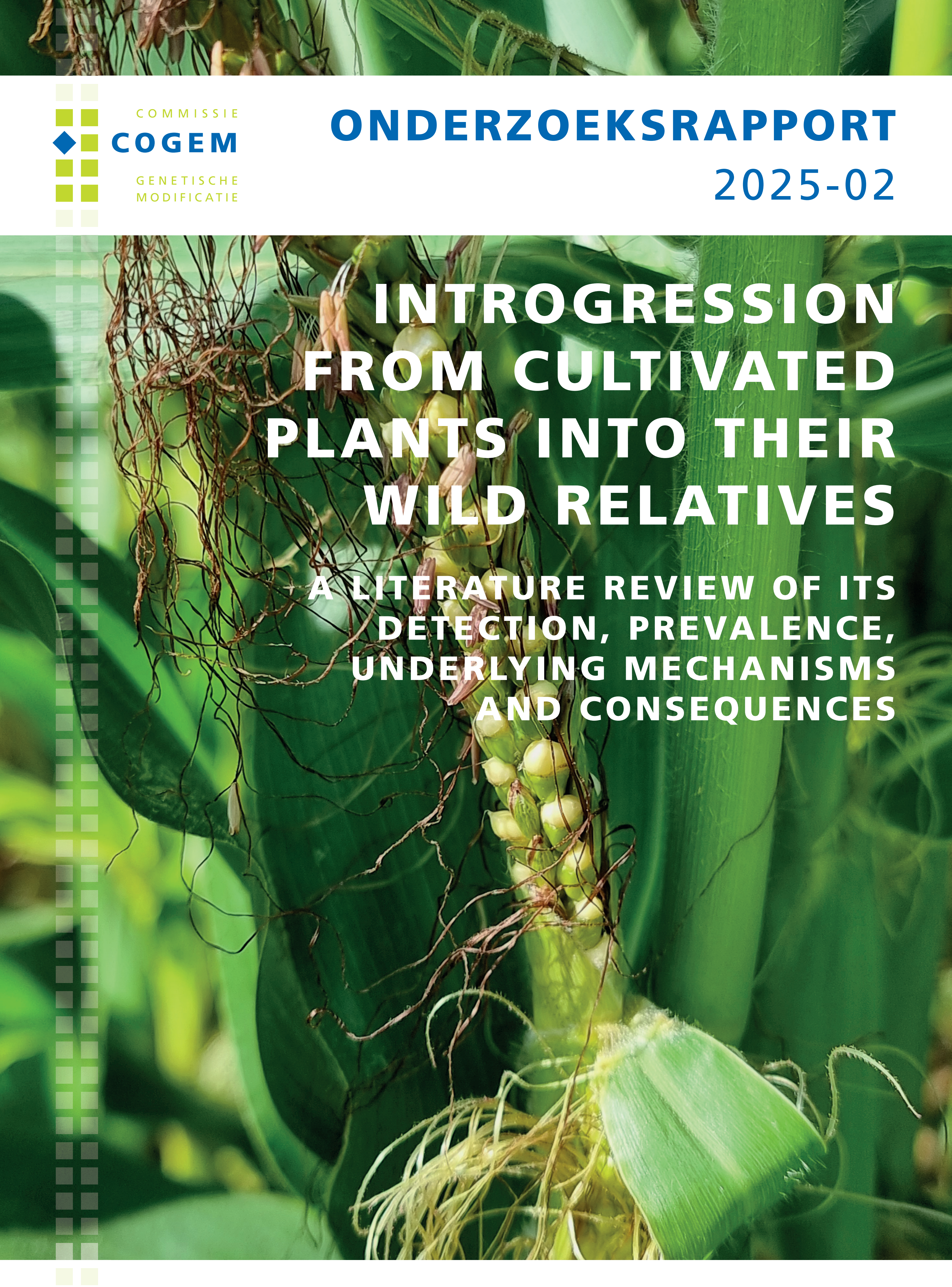
COMMISSIE
COGEM
GENETISCHE
MODIFICATIE

ONDERZOEKSRAPPORT
2025-02



INTROGRESSION FROM CULTIVATED PLANTS INTO THEIR WILD RELATIVES

A LITERATURE REVIEW OF ITS DETECTION, PREVALENCE, UNDERLYING MECHANISMS AND CONSEQUENCES

ONAFHANKELIJKE WETENSCHAPPELIJKE ADVISERING EN SIGNALERING VAN DE RIJKSOVERHEID

Introgression from cultivated plants into their wild relatives

A literature review of its detection, prevalence, underlying mechanisms and consequences

December 2024

Margret Veltman (IRD)
Cécile Berthouly-Salazar (IRD)
Philippe Cubry (IRD)
Muriel Gros-Balthazard (IRD)
Anne-Céline Thuillet (IRD)
Yves Vigouroux (IRD)

French National Research Institute for Sustainable Development
Institut de Recherche pour le Développement
FRANCE

Disclaimer

Cover photo

Close-up of a hybrid teosinte ear, found in a maize field in France.
Photo courtesy of Valérie Le Corre – INRAe.

Acknowledgements

The authors would like to thank everyone who attended the workshop "From hybridisation to introgression: detection and evolutionary consequences", which was held on 25 June 2024 in Montpellier. This workshop helped to shape some of the insights and discussion points that emerged during literature review, and subsequently found their way into this report. In alphabetical order, our sincere gratitude goes to: Concetta Burgarella, Yann Bourgeois, Peter Civáň, Amandine Cornille, Laurène Gay, Jean-Christophe Glaszmann, François Monnet, Miguel de Navascués, Marine Salson and Rémi Tournebize. We would specifically like to thank the workshop speakers for sharing their work on this topic, and all other participants for sharing their collective knowledge and expertise in the discussions that followed.

Voorwoord

Uitkruising van genetisch gemodificeerde (gg-)planten naar wilde verwanten, en daarmee de verspreiding van transgenen in de natuur, is een van de hoofdelementen in de milieurisicobeoordeling van experimenten met gg-planten en gewassen. Met de mogelijke introductie van gewassen die zijn ontwikkeld met nieuwe genomische technieken (NGT) in het milieu is het daarom essentieel om te onderzoeken of introgressie van cultuurplanten naar wilde verwanten plaatsvindt, hoe dit kan worden gedetecteerd, welke factoren hieraan bijdragen, op welke schaal dit gebeurt en wat de mogelijke consequenties zijn voor wilde populaties. Het meest recente literatuuronderzoek naar introgressie dat is uitgevoerd in opdracht van de COGEM stamt uit 2003. Dit rapport is het resultaat van een nieuw literatuuronderzoek (met publicaties uit de periode 2010–2024) en had als doel bovenstaande vragen te beantwoorden en fundamentele informatie te verzamelen om zo risicoanalyses te verbeteren en beter te onderbouwen.

Eerder liet de COGEM onderzoek uitvoeren naar adventieve Koolzaadpopulaties in Nederland, de vorming van hybriden met de wilde verwant Raapzaad, en de aanwezigheid van gg-Koolzaad of hybriden in het milieu. Uit dit onderzoek bleek dat adventieve Koolzaadpopulaties inderdaad voorkomen langs transportroutes, maar het altijd kleine populaties betreft die na enkele jaren verdwijnen. In het materiaal werden geen gg-events gedetecteerd.

Het hier gepresenteerde onderzoek is breder van opzet en richt zich op alle genera en soorten waar in de literatuur onderzoek naar is gedaan naar introgressie van cultuurgewassen naar wilde verwanten. Daarnaast is een overzicht is toegevoegd van het voorkomen van Crop Wild Relatives (CWR)– wilde planten die nauw verwant zijn aan gedomesticeerde planten of gewassen – in Nederland en/of Europa. In dit overzicht is tevens aangegeven of er voor deze soorten/gewassen NGT-onderzoek en ontwikkeling plaatsvindt in Nederland.

Het onderzoek is uitgevoerd door Margret Veltman, werkzaam bij het ‘French National Institute for Sustainable Development’(IRD) in Montpellier. Het rapport is het resultaat van een literatuuronderzoek naar reviews over recente methodologische ontwikkelingen binnen het introgressie-onderzoek, aangevuld met gewasspecifieke studies die in de afgelopen 15 jaar zijn gepubliceerd. Op 25 juni 2024 is in Montpellier een workshop gehouden met experts in gewaspopulatiegenetica uit verschillende instituten, om de theoretische criteria voor claims van introgressie van gewassen naar wilde verwanten scherper te definiëren. Verder hebben drie bijeenkomsten van de uitvoerder met de COGEM-begeleidings-commissie plaatsgevonden, waar ideeën voor opzet en aanpak zijn besproken en de voortgang is gepresenteerd.

Het onderzoek geeft een uitgebreid overzicht van de technische mogelijkheden om introgressie te detecteren, de parameters die introgressie beïnvloeden en studies die introgressie beschrijven op basis van frequentie in populaties of grootte in genomen. Het rapport concludeert dat introgressie naar wilde verwanten wijdverbreid voorkomt, ongeacht de reproductiebiologie of het ploïdeniveau. Het gegeven dat introgressie van gewassen naar wilde verwanten vrijwel onmogelijk te voorkomen is, gecombineerd met de onvoorspelbare consequenties daarvan, betekent dat het essentieel is om mogelijk schadelijke gevolgen te monitoren en te beheren.

Het uitgevoerde onderzoek is bijzonder waardevol voor de COGEM, omdat het cruciale inzichten en handvatten biedt om toekomstige adviezen te formuleren.

Dr. Ir. Guusje Bonnema
Voorzitter van de begeleidingscommissie

Contents

Summary

Genetic introgression between crops and their wild relatives is a common phenomenon. While the majority of research effort goes to studying the realised and potential contributions of crop wild relatives to the domestication and adaptation of cultivated plants, the implications of gene flow in the other direction are less well understood. Previous studies and reviews on the topic have often been framed in the context of managing the risks associated with the release of genetically modified organisms (GMOs) and possible 'transgene escape' into the environment. The development of New Genomic Techniques (NGTs) and their possible exemption from traditional GMO regulations calls for a renewal of attention for this understudied research topic.

This report reviews the latest developments in the study of introgression from cultivated plants to their wild relatives, screening peer-reviewed scientific literature published in the last fifteen years. Complementing a systematic search of scientific literature collected from multiple databases with manually selected articles on species and methods of interest, we collect and report the evidence of crop-to-wild introgression for 25 cultivated plant species (representing 23 genera and 13 families) that were the subject of 53 independent studies. The findings are contextualised with broader insights from previously published reviews and perspectives to answer the questions: 1) What methods are used to detect introgression?, 2) What is the prevalence of crop-to-wild introgression in different cultivated plant species?, 3) What factors determine the likelihood of crop-to-wild introgression?, 4) What are the consequences of crop-to-wild introgression for crop wild relatives?, and 5) How can the (negative) consequences of crop-to-wild introgression be mitigated?

Analysis of the literature showed a surprising breadth of sequencing methods and analytical approaches to study crop-to-wild gene flow. Rooted in population genetic theory, the analysis of introgression requires comparison of sequencing data with statistical methods to detect gene flow between populations or evolutionary lineages. These methods have advanced significantly in the past fifteen years, but many published studies still rely on more traditional genotyping methods and statistical models for population genetic inference. Methodological choices are impacted by trade-offs between sequencing budget and coverage, computational power and resource requirements, and between sampling more lineages versus more representatives per lineage. Methodological difficulties in distinguishing introgression signals from confounding factors (such as population structure and selection) and methodological diversity in sampling design and metrics used, preclude easy generalisation and comparison of these results.

Despite these limitations, studies offered diverse information on the occurrence of introgression, the percentage of the population affected, the genomic extent of introgression and the direction and strength of gene flow events. Global ancestry methods were frequently used to quantify the number of admixed individuals in a population, whereas local ancestry methods and *f*-statistics were commonly used to identify introgressed regions. Population graphs and phylogenetic networks were used to infer the number, timing, direction and strength of gene flow. Other studies used Bayesian modelling to quantify historical and recent migration rates between populations. All these numbers varied widely within and between studies, but a shared insight is that crop-to-wild gene flow seems pervasive, being detected between crops and their close and more distant relatives alike, regardless of mating system and ploidy differences.

Theory on reproductive isolation roughly divides factors influencing gene flow in pre- and post-zygotic mechanisms, with pre-zygotic barriers such as geographic and phenological isolation posing stronger constraints than post-zygotic barriers impacting hybrid fitness, especially in annual species. When hybridisation occurs, the fate of introgressed regions in the genome is influenced by a complex interplay between demography, selection and recombination. This unfolds over multiple generations of backcrossing, with earlier stages generally marked by strong negative selection on long introgression tracts with linked deleterious alleles, and later stages by persistence of (weakly) beneficial alleles in smaller introgression tracts. Which regions are ultimately purged or retained depends on the genomic composition of the founder individuals, the recombination landscape of the genome, and the environment that exerts selective pressures on the hybrid individuals and backcrosses.

As a result of these interlinked factors, the consequences of introgression are varied, and depending on the donor and recipient population they can range from genetic assimilation of the parent populations; outbreeding depression, populating decline and collapse (and possibly extinction), to increased diversity and genetic rescue of fragmented populations; and strong selective sweeps on specific introgressed alleles known as adaptive introgression. The most commonly reported consequences were adaptive introgression, leading to the acquisition of weedy traits in several crop wild relatives (supported by identification of introgressed crop alleles underlying relevant phenotypic traits), and genetic assimilation (often inferred from high levels of crop genetic ancestry in wild individuals). Genetic rescue of crop wild relatives by introgression of crop alleles is an intriguing possibility, but only suggested in one study.

In conclusion, crop-to-wild introgression is common across many different types of crops, but its consequences are hard to predict. Small and fragmented populations are more susceptible to genetic or demographic swamping by crop-to-wild gene flow, while species undergoing range expansion or human-mediated introductions to new areas might benefit from introgression of locally adapted crop alleles and quickly become weedy or invasive. These risks pertain to crop-to-wild gene flow regardless of whether the crop is a traditional cultivar, GMO, or engineered with NGTs - although genetic modification and NGTs can accelerate the development of traits controlled by few loci that are susceptible to rapid introgression when adaptive in wild or weedy populations. Risks that are specific to NGT plants therefore depend on the engineered traits and its genomic architecture. Unfortunately, currently available mitigation options are not fool proof and achieve partial reproductive barriers at best.

The near-impossibility to prevent gene flow and its unpredictable consequences means that monitoring is an essential strategy to track and control possible adverse consequences of crop-to-wild introgression. While introgression is a multi-dimensional phenomenon that can be studied in a myriad of ways, monitoring crop-to-wild introgression in areas where it hasn't occurred yet will benefit from dense population sampling and might not require genome-wide coverage. An exception concerns tracking the introgression of specific engineered crop alleles, which would require targeted sequencing or full genome data. Allocating monitoring resources requires an evaluation of the social acceptability of (NGT) crop-to-wild gene flow and the possibly unforeseen consequences it might have for local species. As different actors in society engage in this discussion, scientific advances will continue to enable ever more fine-grained studies of introgression and its dynamics in different organisms. Prioritisation of cultivated plant species for future research and risk assessments can benefit from a list of priority taxa identified in this review, including genera rich in crop wild relatives occurring in Europe and the Netherlands, and those with cultivated species currently targeted for improvement with New Genomic Techniques.

Glossary

Adaptation	The dynamic evolutionary process by which biological organisms develop characteristics that allow them to survive and reproduce within their environments.[1]
Admixture	Admixture occurs when multiple divergent genetic lineages come into gene flow contact and interbreed.[2]
Allopatry	Allopatry, meaning ‘in another place’, describes a population or species that is physically isolated from other similar groups by an extrinsic barrier to dispersal.[3]
Allopolyploid	A polyploid cell or organism in which the several sets of chromosomes originate from more than one species, as in an intraspecific hybrid. Different from autopolyploid, which possess multiple chromosome sets derived from a single taxon.[1]
Backcrossing	Backcrossing is a type of recurrent hybridization in which a gene for a desirable trait is added to a desirable variety.[3]
Crop wild relative	Crop wild relatives (CWR) are wild plant species that are genetically related to crops, but have not been domesticated.[3]
De-domestication	De-domestication or feralization is an evolutionary process by which domesticated crops or livestock reacquire components of wild-like traits and escape from intensive management by humans to form independent reproducing populations (Wu et al., 2021).
Domestication	A coevolutionary process that arises from a mutualism, in which one species (the domesticator) constructs an environment where it actively manages both the survival and reproduction of another species (the domesticate) in order to provide the former with resources and/or services (Purugganan, 2022).
Domestication syndrome	The characteristic collection of phenotypic traits associated with the genetic change to a domesticated form of an organism from a wild progenitor form (Allaby, 2014).
Effective population size	The number of individuals that effectively participates in producing the next generation. It is typically smaller than the total census size and plays a significant role in influencing evolutionary processes.[4]
Feralisation	See "De-domestication".
Gene flow	The transfer of genetic variation from one population to another, by any available means, e.g., by sexual reproduction, horizontal gene transfer, or retroviral integration.[1]
Genetic drift	A change in the frequency with which an existing allele occurs in a population due to random variation in the distribution of alleles from one generation to the next.[1]
Genetic engineering	A process that uses laboratory-based technologies to alter the DNA makeup of an organism.[5]

Genetic load	Any reduction in the mean fitness of a population owing to the existence of one or more genotypes with lower fitness than that of the most fit genotype.[1]
GMO	A genetically modified organism (GMO) is a plant, animal or microbe in which one or more changes have been made to the genome, typically using high-tech genetic engineering, in an attempt to alter the characteristics of an organism.[5]
Homoplasy	Homoplasy is the term used to describe two characters whose similarity derives from convergence as opposed to common descent.[2]
Hybridisation	The process by which a hybrid organism is produced from two parents of different genera, species, breeds, or varieties.[1]
homoploid ~	Hybridisation between two organisms possessing the same ploidy level.
interploid ~	Hybridisation between two organisms possessing a different ploidy level.
Introgression	The movement of a gene from the gene pool of one population or species into that of another population by the repeated backcrossing of hybrids of the two populations with one of the parent populations.[1]
Invasive species	An invasive species is an introduced species (also called “nonindigenous” or “non-native”) that adversely affects the new habitat ecologically or economically. Invasive species may be microbes, plants, or animals and may disrupt the new area by dominating particular habitats because of loss of natural biological controls present in its native range, such as predators or diseases.[4]
Linkage	Linkage, as related to genetics and genomics, refers to the closeness of genes or other DNA sequences to one another on the same chromosome. The closer two genes or sequences are to each other on a chromosome, the greater the probability that they will be inherited together.[5]
Linkage disequilibrium	Linkage Disequilibrium (LD) is a term coined to describe the non-random association between alleles of genetic loci that lie in proximity to each other.[3]
NGTs	New Genomic Techniques (NGTs) are methods for creating targeted mutations (mutagenesis) in the genome of living organisms. Also called gene editing techniques. In the EU, NGTs are defined as "techniques that are able to alter the genetic material of an organism, developed after the publication of EU Directive 2001/18/EC" (Parisi & Rodríguez-Cerezo, 2021)
Polyploid	Having more than two homologous copies of each chromosome.[1]
Recombination	Genetic recombination is the physical breakage, exchange, and rejoining of two DNA molecules.[2]
Selection	The non-random differential survival or reproduction of classes of phenotypically different entities.[1]

Selective sweep	The process by which strong positive selection of a new and beneficial mutation within a population causes the mutation to reach fixation so quickly that nearby linked DNA sequences also become fixed via genetic hitchhiking, thereby reducing or eliminating the genetic variation of nearby loci within the population.[1]
Sympatry	Sympatry is the term used to describe populations, varieties, or species that occur in the same place at the same time.[3]
Transgene	Property of an organism or cell whose genome has been altered by the introduction of one or more foreign DNA sequences from another species by artificial means.[5]

Consulted reference works (glossaries, encyclopaedia entries, and scientific topic collections):

[1] https://en.wikipedia.org/wiki/Glossary_of_genetics_and_evolutionary_biology
[2] https://www.sciencedirect.com/topics/biochemistry-genetics-and-molecular-biology
[3] https://www.sciencedirect.com/topics/agricultural-and-biological-sciences
[4] https://www.sciencedirect.com/topics/earth-and-planetary-sciences/
[5] https://www.genome.gov/genetics-glossary

1 Introduction

1.1 Background

Genetic modification and New Genomic Techniques

Introgression between crops and wild relatives is a common phenomenon (Ellstrand et al., 2013). Much research effort has focused on the role of introgression in the evolution of domesticated species, with evidence of adaptive introgression between crop varieties or from wild relatives into crops in numerous species (Janzen et al., 2019). However, less attention has so far been paid to the possible consequences of introgression from crops into their wild counterparts, despite concerns about possible adverse effects such as weediness and loss of genetic diversity (Feurtey et al., 2017; Le Corre et al., 2020).

An exception to this is the debate surrounding the regulation of genetically modified organisms (GMO), which began in the 1970s (Berg et al., 1974; Singer & Berg, 1976) and was revived around the turn of this century (Jenczewski et al., 2003), spurring considerable interest in the possibility of releasing transgenes and otherwise genetically modified DNA into natural populations. Early studies investigating (genetically modified) crop-to-wild gene flow were conducted in a range of species, including beet, cotton, rapeseed, maize, radish and rice (Ellstrand, 2003b; Groot et al., 2003). In addition to these species-specific observations, a large-scale survey of evidence from experimental crosses suggests that at least half of the crops cultivated in temperate regions are reproductively compatible with one or more of their wild relatives (Armstrong et al., 2005). Strict regulations in the EU, requiring pre-market authorisation based on a prior risk assessment, traceability and labelling of GMOs (Bruetschy, 2019), subsequently limited the cultivation and import of GMOs, abating concerns about the escape and establishment of transgenic individuals; although accidental releases into the environment are occasionally reported (Rostoks et al., 2019).

The development of New Genomic Techniques (NGTs), including gene-editing tools such as CRISPR/Cas-9, has renewed concerns over the risk of introducing genetically engineered material into natural populations (Ahmad et al., 2023). Operating on the precautionary principle, NGTs were initially subject to the same strict regulations as traditional GMOs (Stokstad, 2018). Recent reports have called for revision of the European GMO laws and a more differentiated approach to their regulation (Stokstad, 2023). This would exempt gene-edited plants that are considered equivalent to what could be achieved with conventional breeding from compliance with GMO regulations, and require a case-by-case assessment of gene-edited plants that are not considered equivalent (Dima et al., 2023).

New data and tools to study introgression

In order to make an informed risk assessments of introgression from GMOs and gene-edited plants into wild populations, baselines need to be established for the prevalence, mechanism and effects of introgression from crop to wild populations (Sanvido et al., 2012). These baselines are currently missing or sparsely documented for a limited number of species (Andersson & de Vicente, 2010; Ellstrand et al., 2013). However, the last decades have generated a wealth of literature on the detection and extent of introgression in a variety of study systems, boosted by an exponential growth in the availability of DNA sequencing data and the number of tools available for their analysis (Edelman & Mallet, 2021).

Specifically, the genomics revolution, triggered by the development of next generation sequencing technologies, has opened up the possibility to sequence entire genomes of almost any organism at high quality and increasingly low cost. Since the publication of the human genome (Lander et al., 2001), reference genomes have been released for numerous plant species, including many crops (Kress et al., 2022). At the same time, computational and statistical advances in population genomics have generated increasingly sophisticated methods that can analyse vast quantities of data (Luikart et al., 2019). The combined availability of high-quality reference genomes and analytical techniques has enabled the precise detection of signatures of gene flow along the genome, both in recent and in deep time (L. Chen et al., 2020; Le Corre et al., 2020). This means that the study of introgression has now, for a few years, been possible at an unprecedented scale and at unprecedented genomic resolution, catalysing discoveries of gene flow in a wide range of biological systems, and in different spatial and temporal contexts.

Knowledge advances and gaps

The study of crops and their wild relatives has benefited from this increasing resource availability. While much effort in the development of genomic data and tools is motivated by (commercial) interests in breeding, there have been spill-over effects into more fundamental studies. These have focused on the evolutionary history and origins of crops, as well as their relationships with wild relatives (Meyer & Purugganan, 2013). Recent years have seen the use and implementation of various statistical tools for detecting introgression in diverse crop systems, including haplotype structures (Vi et al., 2023), *D*- and *f*-statistics (Le Corre et al., 2020) and admixture graphs (Brandenburg et al., 2017). Most of these methods were initially developed in the context of (ancient) human genetics, driven by an interest in early human history and migration (Green et al., 2010; Patterson et al., 2012). But despite some challenges in their application to organisms with different ploidy levels and reproductive systems, these tools are now widely applied and adapted to other species, including wild and domesticated plants (Burgarella et al., 2019).

Population genomics studies are therefore increasingly illuminating the molecular signatures and extent of introgression across study systems and evolutionary time scales (e.g., Scalabrin et al., 2024; Wang et al., 2017). A synthesis of these studies could reveal new insights concerning the biological conditions restricting or promoting crop-to-wild gene flow and its potential (mal-)adaptive effects in wild populations. Yet, the actual effects of introgression are less easily uncovered through genetic means and, without observational or experimental data, are often surrounded with a degree of uncertainty. Estimates of fitness, population census size, and phenotypic traits typically require experimental approaches or extensive surveys; in the absence of systematic efforts to monitor wild populations, many of these observations are likely to remain anecdotal (Feurtey et al., 2017). In addition to empirical studies, theoretical predictions can be made regarding the factors that influence the likelihood and effects of introgression, such as mating system, generation time, seed dispersal mechanism, propagation method, genetic incompatibilities and other mechanisms of reproductive isolation (Tenaillon et al., 2023). These will carry different implications for the risk of introgression in systems where these factors vary, such as annuals versus perennials (Gaut et al., 2015), but haven't been systematically placed in the context of crop-to-wild introgression to date.

1.2 Objectives

This review aims to synthesise recent methodological and empirical advances made in the study of crop-to-wild introgression, as well as knowledge gaps remaining. Through a comprehensive assessment of the available literature documenting crop-to-wild introgression with the use of state-of-the-art genomic techniques, we summarise what is currently known about the prevalence of crop-to-wild introgression, the factors that influence its occurrence, and its possible effects on crop wild relatives (CWR), through the lens of population genomics analysis. We do this by systematically screening the available literature in order to answer the following research questions:

1. How is introgression detected and monitored?
2. How common is crop-to-wild introgression in different crop species?
3. What factors and mechanisms determine the possibility and extent of crop-to-wild introgression?
4. What consequences does crop-to-wild introgression have for natural populations of crop wild relatives?
5. How can the extent of crop-to-wild introgression and any adverse consequences it might have for wild populations be mitigated?

To answer these questions, this report will first chart the methodological developments that have enabled ever more fine-grained studies of introgression, as well as their promises and their pitfalls (Methods). Subsequently, it will cover how often introgression is detected in the different crop systems surveyed, focusing on different metrics that are used to quantify introgression (Prevalence). These findings are then discussed in light of the factors that are known to influence the occurrence and extent of gene flow (Factors), and the possible outcomes of introgression under different conditions (Consequences). Finally, tying together knowledge of the detection methods, prevalence, causes and consequences of introgression from crops to wild relatives, we will discuss how any risks for crop wild relatives associated with the use of NGTs in European agriculture could be monitored and mitigated (Discussion).

The answers to the research questions outlined here will hopefully contribute to a better understanding of the factors influencing the occurrence of introgression, the processes that lead to its persistence, and the possible consequences for crop wild relative diversity and adaptation, thereby improving the knowledge base for future risk assessments of genetically modified and gene-edited crops.

1.3 Terminology

This report makes use of some technical concepts that require definition. The most important interrelated concepts regarding introgression and crop wild relatives are disambiguated below. Brief definitions of these and other concepts throughout the report are given in the Glossary.

Gene flow, hybridisation, introgression and admixture

Different technical terms are used to describe (evidence of) the exchange of genetic material between populations. Gene flow is the movement of alleles between two populations or species, and can occur in discrete (time-bound) events called "pulses", or in a continuous flow over time (often measured in terms of a "migration rate"). Where gene flow describes the process or pressure behind the incorporation of foreign alleles into a population, either through seed dispersal or through pollen dispersal, introgression refers to the (permanent) incorporation of genomic segments from a donor population into a recipient population through hybridisation and backcrossing, and hence describes the resulting genomic pattern. In this definition, hybridisation is considered the first step in the process of introgression, where crossing between two genetically distinct populations or species results in a first generation of hybrids. The resulting hybrid offspring will contain roughly equal proportions of both parental genomes. Backcrossing involves consecutive crosses between these hybrids and one or more parental populations over multiple generations, diluting or increasing the proportion of one parental genome relative to the other and leading to distinct patterns of introgression. While introgression is per definition localised, signs of introgression be detected either globally (across the whole genome) or locally (in specific regions of the genome). In the latter case, the regions that are introgressed are referred to as "introgression tracts", or "introgression blocks". A related concept is "admixture", which occurs when divergent lineages come into contact and interbreed, resulting in populations of mixed ancestry. Whereas gene flow can occur at low levels and the resulting patterns of introgression often imply a direction from a donor to a recipient population, admixture presumes extensive bi-directional gene flow, leading to more similar proportions of ancestry throughout the genome.

Crops, wild progenitors, wild and weedy relatives, and feralisation

Gene flow or admixture between diverged lineages can happen at different points along the speciation continuum (and at different levels of divergence), from recently separated populations to species that belong to completely different genera. Species boundaries between crops and their wild relatives are notoriously fuzzy, with different species names (often, not always) given to the domesticated species and its wild progenitor, despite at most a few thousand years of evolutionary separation. Divergence times between a crop and its relatives other than the wild progenitor are usually much older, and of the same age as the separation between the relative in question and the progenitor species. For the purpose of this review, we will therefore make no distinction between intra- and inter-specific gene flow, but we will consider whether gene flow has occurred between a crop species and its progenitor or more distant relatives. These relatives may occur in weedy forms and be referred to in the literature as "weedy relatives". The genetic pathways to weediness are varied and may involve (repeated) gene flow with the cultivated species, adding to the blurry boundaries between cultivated and wild genotypes. Escape of cultivated plants and subsequent adaptation to the wild (also known as de-domestication or feralisation) can likewise give rise to phenotypes that are intermediate between the crop and the wildtype, and lead to the emergence of weedy forms. These distinctions are not always clear, but for the purposes of this review we consider a wild or weedy relative any wild population capable of sustaining itself without human intervention.

2 Approach

To answer the questions outlined above, we adopted a two-pronged approach to this review. We first conducted a preliminary literature review drawing from previous reviews, perspectives and reports the topic (e.g., Ellstrand, 2014; Ellstrand et al., 2013; Groot et al., 2003; Hibbins & Hahn, 2022), and the references cited therein, to provide background knowledge on the research questions stemming from current scientific consensus and theory. We subsequently searched the scientific literature for peer-reviewed articles mentioning crop-to-wild or crop-to-weed gene flow (Box 1) in their title, abstract or keywords, in order to find recent empirical examples of crop-to-wild introgression. This literature was screened for eligibility and scored to assess the methodologies used (*RQ1*), quantify the evidence base for introgression in different groups of crops (*RQ2*) and shed further light on the causes (*RQ3*) and consequences (*RQ4*) of introgression in specific study systems. With these results, we then re-evaluated the preliminary answers to *RQ1* and *RQ5* by comparing how introgression is (or is not) detected and managed in practice versus how it *could be* detected or managed. Below we describe the parameters and processing of this second search.

2.1 Scope

Since we were primarily interested in gene flow from crops into wild populations, but publications investigating the reverse direction (from wild populations into crops) are decidedly more numerous, we restricted our search to papers explicitly addressing gene flow in the crop-to-wild direction. We opted for this conservative approach at the risk of missing some of the available evidence reported in papers that do not explicitly refer to crop-to-wild introgression in the abstract, in order to keep the volume of literature manageable and on point. Our search thus prioritised specificity over sensitivity, and likely does not capture all the available empirical evidence that exists in the scientific literature. Supplementary studies prioritising sensitivity should follow up on specific species of interest that are of special interest for future risk assessments (see Priority taxa). Since our focus on directional crop-to-wild gene flow drastically reduced the number of articles to review relative to all articles mentioning gene flow between crops and wild relatives regardless of its direction, we decided not to restrict our scope geographically, but consider all evidence of crop-to-wild gene flow globally. This review therefore also makes reference to species that are not cultivated in Europe or even in temperate regions. We hope that this global overview of crop-to-wild introgression can serve as a starting point for more thorough, regionally oriented surveys.

2.2 Priority taxa

In order to place the current results in the context of future NGT risk assessments, we drafted a list of all crop wild relatives that occur in the Netherlands (van Treuren et al., 2017, 2020) and/or are priorities for conservation in the EU (Rubio Teso et al., 2020), as well as all species that were the subject of active research and development projects using NGTs as of January 2024 (Anses, 2024). Since these different sources use different nomenclatures, we cross-checked all taxon names and resolved any consistencies. We did this by updating species and genus names with the R package 'taxize' (Chamberlain & Szöcs, 2013), prior to tabulating the species counts per genus. We subsequently matched the updated names to taxon identifiers in the World Checklist of Vascular Plants (WCVP) to obtain higher level taxonomic classifications.

Box 1. Search terms used to identify potential studies be included in the review

Figure 4 shows the search criteria that were used to collect literature and formed the basis of specific search terms. These search terms were adapted to the Web of Science and Scopus databases in the following way.

Scopus:	*(TITLE-ABS-KEY (({crop-to-wild} OR {crop-weed} OR "crop-to-weed" OR "crop-to-wild") AND ("introgress*" OR "gene flow" OR "hybridi*" OR "admix*" OR outcross*)) AND LANGUAGE (english)) AND PUBYEAR > 2009*
Web of Science:	*PY=(2010-2024) AND LA=(English) AND TS=(("crop-to-wild" OR ("crop-to-wild" NOT "crop wild relative*") OR "crop-to-weed" OR "crop-weed") AND ("introgress*" OR "gene flow" OR "hybrid*" OR "outcross*"))*

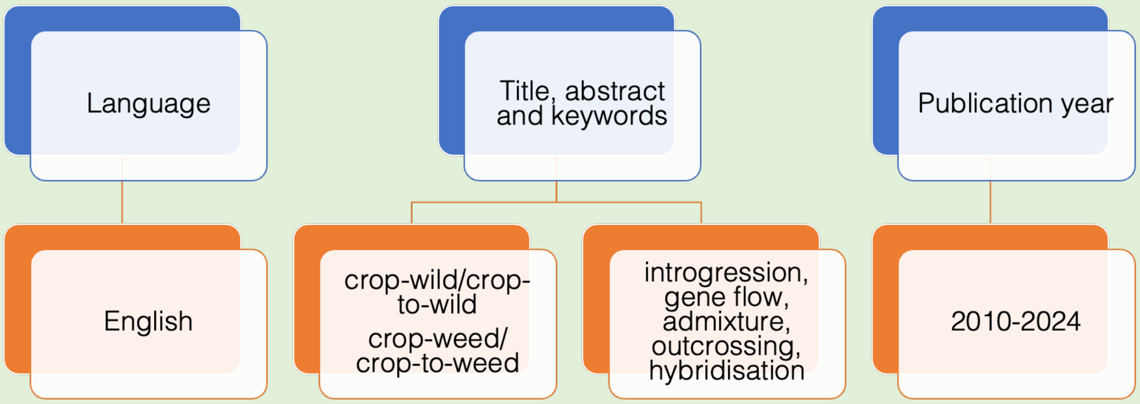


Figure 4. Schematic overview of search terms used to collect literature.

By intersecting the genera and families to which these key species belong, we were able to define target taxa of high priority for risk assessment in the Netherlands and in the EU, that would merit more thorough review[1] (Annex A). We used this overview to supplement our literature collection, where possible, with additional papers on genera present in all three lists.

2.3 Literature collection

Our literature collection and screening approach was guided by the PRISMA-RR protocol for rapid reviews (Stevens et al., 2024). Since the objective of this research was to synthesise the latest scientific insights, we limited our search to the past 15 years (2010-2024). This time period follows the emergence of next-generation sequencing for population genetic studies, tracking its spread to increasing numbers of model and non-model plants. It also runs parallel

---

[1] R&D projects are conducted on cultivated species. By selecting the genera to which these species belong, we are aware that we might overlook crop wild relatives that are not congeneric with the cultivated species. However, the number of species with geographically relevant CWR in other genera is limited to a small number of crops, most notably wheat (nine genera in the EU, of which two in the Netherlands), rapeseed (eight genera in the EU, of which six in the Netherlands), and strawberry (two genera in the EU and in the Netherlands).

with the development of *D*-statistics and many of its derivatives. These were initially developed and applied to the study of gene flow between modern humans and Neanderthals (Durand et al., 2011; Green et al., 2010; Patterson et al., 2012) and have subsequently been adopted by other research communities applying them to different model systems including wild plants and crops (Burgarella et al., 2019; Janzen et al., 2019; Suarez-Gonzalez et al., 2018). We restricted our search to English language articles published in peer-reviewed journals indexed in Scopus and/or Web of Science, mentioning crop-to-wild introgression (Box 1).

After an assessment of the commonly used methods and reported results, we manually supplemented our search with additional articles to increase the available information on relevant crop systems or methods not extensively covered in the systematic survey, specifically the *D*-statistic developed by Green et al. (2010) and the 12 high priority genera defined in Table A1. We subjected the articles that were found through the initial and supplementary search to the same eligibility criteria and data extraction protocol. All search queries were adapted to the Scopus and Web of Science databases, and results were exported in RIS format (a standardised bibliographic file format developed by Research Information Systems, to facilitate exchange between different citation management software), for importation in the citation manager Zotero.

2.4 Eligibility and data extraction

Following import and deduplication in Zotero, deduplicated records were exported as RIS and uploaded to Rayyan (Ouzzani et al., 2016), a web-based tool for conducting systematic reviews. We used Rayyan for title and abstract screening, keeping only articles that reported 1) using genetic data, 2) obtained from populations occurring under natural (i.e., non-experimental) conditions, 3) to study introgression, 4) from crops into wild (or weedy) relatives. Despite narrowing our search to "crop-to-wild" and "crop-to-weedy" gene flow, we still encountered some articles reporting on gene flow in the other direction, which shows that even our selective approach yielded articles outside the scope of the immediate research question. Apart from failing to meet the basic inclusion criteria of being an English language, peer-reviewed journal article published since 2010, studies were therefore discarded if they did not explicitly study crop-to-wild introgression, or if they studied crop-to-wild introgression, but without using genetic data or by characterising it only under experimental conditions[2] (Figure 1). For the remaining studies, we proceeded to retrieve the full text to assess their eligibility. Based on the full text we discarded studies for the same reasons as above, while applying an additional criterion: if the methods and results did not describe how introgression was detected or quantified (i.e., lacking an operational definition of introgression), the study was excluded (Figure 1).

For studies that were included, we noted the study system (crop species, wild relative species, geographic location), study design (number of individuals and populations sampled, sequencing technique used and type and number of markers used for population genetic inference), analytical method (tool or statistic used) and main findings (the unit and quantity of introgression detected, where applicable), depending on the tool or statistic. More articles were discarded during this stage, either because they were found to be lacking detail in their

---

[2] Even though experimental studies can offer interesting insights into the determinants and mechanisms of introgression, the central focus of this review was to collect evidence of the occurrence of crop-to-wild introgression in nature, and hence experimental studies were not prioritised.

reporting, or because they reported conclusions that that did not follow from the results. The numbers of articles found, screened and assessed, are shown in Figure 1. The data extracted is summarised in Annex B.

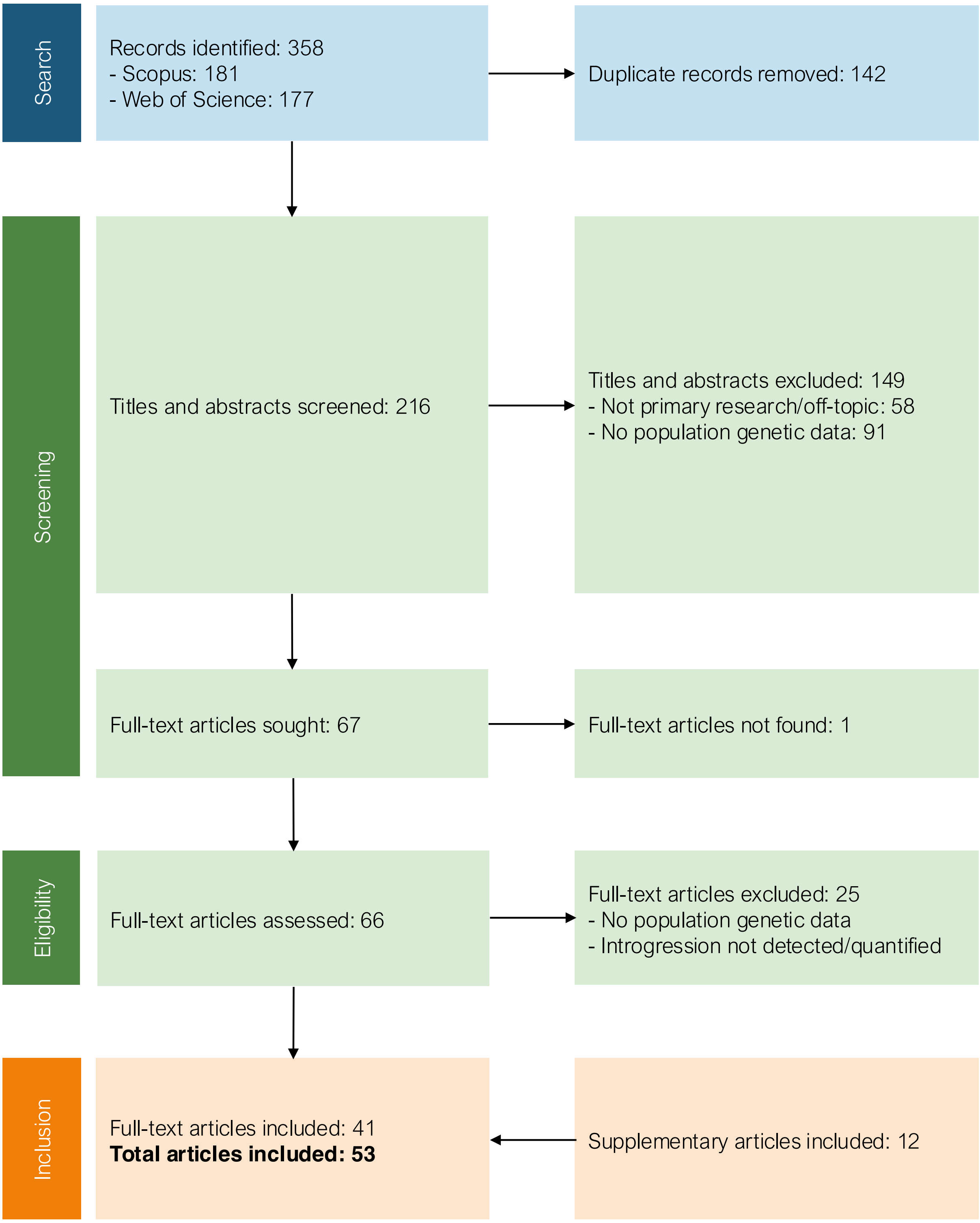


Figure 1. PRISMA-RR flow diagram, showing the number of studies searched, screened, and found eligible for inclusion.

2.5 Overview of the surveyed literature

In total, data on crop-to-wild introgression was extracted from 53 articles, covering a wide range of species, geographical areas, sequencing and analytical methods, and findings. These data are available in tabular format (Table B1), and are narratively described in species-specific summaries in Annex B. In brief, the majority of articles studied grasses and cereals, or fruit and nut-bearing tree crops, with a clear bias towards crops from North America and Europe, and a lesser focus on African, Asian and Australian crops. A surprisingly large number of articles (about half of all surveyed studies) did not make use of Next Generation Sequencing (NGS), but rather relied on older genotyping methods (Figure 2A). This limited the analytical methods available to detect introgression (which are described in more detail in the next section), but allowed for larger sample sizes, ranging in the hundreds or even thousands (Figure 2B). In the studies we selected, introgression was always quantified, but rarely explicitly linked to causal factors or consequences. We therefore discuss the latter in a more general fashion. The next sections describe in detail how the studies we surveyed analysed introgression, the extent of introgression they found, the biological factors that might contribute to these findings and, where reported, the consequences of introgression for crop wild relatives.

3 Methods to detect introgression

3.1 Sequencing, markers and sampling

The surveyed studies made use of a wide diversity of sampling designs, sequencing methods, and analytical tools to detect, characterise and quantify introgression (Table B1). The most commonly used markers were found to be microsatellites, with single nucleotide polymorphisms (SNPs) coming in second (Figure 2A). A minority of studies were based on allozymes, amplified fragment length polymorphisms (AFLPs), or random amplified polymorphic DNA (RAPD), all of which were published more than 10 years ago. Many crops are model systems, which means they are often associated with reference genomes and other genomic resources, making them amenable to genome-wide analyses of evolutionary processes and relationships. Nonetheless, despite the availability and dropping costs of next-generation sequencing methods, microsatellites remained a popular and cost-effective method to screen genetic diversity and population structure for many of the studies reviewed here, with articles using this method published in almost every year (Figure 2A).

Studies based on SNPs tend to be published more recently and are most commonly based on genotype-by-sequencing (GBS) or restriction site-associated DNA (RAD) sequencing. These are reduced representation sequencing strategies relying on the random cut sites of restriction enzymes to amplify and sequence a limited set of loci which should in theory be randomly distributed across the genome. Other reduced representation techniques such as target enrichment and capture (of exomes, selected nuclear genes or other loci) or RNA sequencing are also encountered, as is low pass whole genome sequencing, (where the entire genome is sequenced at a low depth of coverage, also known as "genome skimming"). Reflecting the higher costs and computational resource investment associated with sequencing and bioinformatic analysis of entire genomes, SNPs based on whole genome re-sequencing studies are less commonly encountered in our survey, and emerge in this body of literature relatively recently. The most advanced and complete representations of genomic diversity included here are *de novo* assemblies and pangenome graphs, which are employed in our selection of studies only twice: both in articles on Asian rice, which was one of the first species to have a published reference genome and currently has high quality genome assemblies for multiple (sub)species and relatives within the genus. This is possibly explained by the fact that, at ~400Mb, rice has a relatively small genome compared to other major cereals, such as maize and wheat, facilitating the use of such an approach.

The differences in genotyping or sequencing strategy are reflected in sampling design (Figure 2B). Most microsatellite studies are geographically explicit studies screening large numbers of individuals from multiple populations. On the other hand, broader and deeper coverage of one or multiple genomes is often accompanied by a stronger focus on evolutionary processes, reducing single or few individuals to evolutionary lineages and requiring less dense population sampling, but placing these into a broader phylogenomic or taxonomic context. To some extent, these approaches reflect disciplinary practices and their underlying assumptions and research aims: in conservation science, introgression might be studied with a temporally narrow, and geographically explicit, ecological focus, offering insights that are directly relevant for applications in society; while in evolutionary science, introgression might be studied on deeper time scales to address fundamental questions about the history of a crop species. The articles surveyed here therefore reflect different types of studies produced by slightly different research communities, yet they are unified by using a shared framework of population genetic theory and analysis. All are relevant for GMO or NGT risk assessments, insofar as they provide evidence that introgression can occur or has occurred, albeit possibly at different geographic and temporal scales. The relevance of these disciplinary approaches for crop-to-wild gene flow monitoring in the context of GMO and NGT crops, is elaborated further in the Discussion.

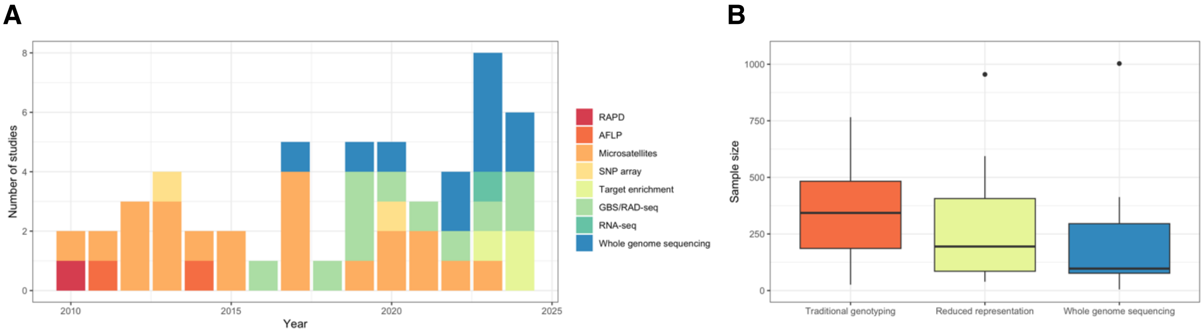


Figure 2. Overview of genotyping methods used in the studies surveyed, and their relationship to sample size. A) Number of studies surveyed per publication year, broken down by genotyping method. B) Differences in sample size distribution for different classes of genotyping methods. Five outliers with > 1200 samples were removed for legibility of the graph. Traditional genotyping = RAPD, AFLP and microsatellites; Reduced representation = Target enrichment, SNP array, GBS/RAD-seq and RNA-seq. Genotyping method significantly impacts sample size (Kruskal-Wallis test, $p<0.05$).

3.2 Analytical methods

Similar to sampling design, the chosen markers and sequencing strategies have implications for the types of analytical methods that can be used to study gene flow. These methods can be broadly categorised into two classes: population genetic methods and phylogenetic methods (Figure 3). Population genetic methods rely on allele frequencies and summary statistics of genetic diversity to assess allele sharing and local deviations in ancestry compared to a genome-wide average. Phylogenetic methods rely on the presumed or inferred relationships between multiple lineages reflecting the "true" or genome-wide genealogy, against which alternative site patterns or gene tree topologies are compared to statistically test whether discordant patterns are caused by gene flow. We will discuss both classes of methods and their applications in the surveyed literature below. In addition to these methods for detecting gene flow, other methods exist for identifying and classifying hybrids, including the use of diagnostic alleles, parentage analysis, and calculation of the hybrid index (HI). These are also occasionally reported in the literature we surveyed, but we will not describe them in detail here. An overview of the encountered statistical approaches and/or software packages implementing them, as well as their scope of application, is given in Table 1.

Population genetic methods

Population genetic methods tend to be based on polymorphism data and require a larger number of individuals to infer population structure and degrees of genetic differentiation and/or gene flow between them. Among these methods are summary statistics, such as pairwise nucleotide diversity (d_{xy}) and pairwise population differentiation (F_{ST}). These can be used to infer the migration rate (for example using the equation $F_{ST}=1/(1+4N_m)$ to calculate the number of migrants per generation N_m), and they can be scanned along the genome to identify regions where the number of nucleotide differences between populations or the level of genetic differentiation significantly deviates from the genome-wide average. An example of this is the detection of transgene flow in Ethiopian sorghum (Adugna & Bekele, 2017). However, it has been argued that such indirect measures of gene flow often rely on unrealistic biological assumptions (such as equal population size, equilibrium between migration and drift and the absence of selection) which make them unsuitable for the inference of migration in natural populations (Whitlock & McCauley, 1999)

In addition to summary statistics, there are model-based inferences of admixture that use allele frequencies to infer population structure and gene flow. One type of inference results in admixture graphs, which models the demographic history of populations by taking the estimated population tree as a basis, and subsequently adding migration edges to this tree to improve the likelihood of the model (Pickrell & Pritchard, 2012). This method has been used in multiple SNP studies to identify migration events from crop to wild populations in diverse systems such as hazelnut (Helmstetter et al., 2020), eggplant (Barchi et al., 2023; Page et al., 2019), maize (Le Corre et al., 2020), olive tree (Zunino et al., 2024) and rice (Qiu et al., 2020). While Approximate Bayesian Computation (ABC) methods such as DIYABC (Cornuet et al., 2008, 2014) can similarly be used to infer population divergence and admixture, they differ from TreeMix in that they allow the user to define more complex demographic scenarios. What makes both models attractive, is that they recognise that population histories are rarely strictly bifurcating and that they allow researchers to test and compare different models to determine the number (and direction) of gene flow events, as well as the strength of gene flow (TreeMix) or proportion of admixture (DIYABC), and the timing of admixture events (Figure 3B).

Another type of inference models the mixing of hypothetical ancestral populations resulting in different ancestry proportions in individual genomes. These ancestry proportions are presumed to arise after secondary contact between a number of discrete ancestral populations (following a period of isolation), and can be used to identify admixed individuals based on the relative proportions of the genome (or available amount of genetic data) that can be attributed to these ancestral populations (Figure 3A). There is now a range of such model-based methods for ancestry inference, including STRUCTURE (Pritchard et al., 2000), InStruct (Gao et al., 2007), ADMIXTURE (Alexander et al., 2009), and fastSTRUCTURE (Raj et al., 2014), which have individually been cited thousands of times and whose popularity for identifying admixture is confirmed by the fact that they compromise a large portion of the evidence included in our survey, representing over half the studies (Table B1).

However, violations of model assumptions (similar to the summary statistics described above) and imbalanced sampling can lead to artefacts, causing spurious results. In addition, even when model assumptions are met, the results might represent historical gene flow and do not reveal much about the timing or mode of the gene flow events; hence, they should be interpreted with caution (Lawson et al., 2018). Finally, these methods do not provide a statistical test of specific hypotheses of gene flow, and rely instead on the definition of arbitrary thresholds of ancestry proportions to determine whether an individual is "admixed". In many studies, this threshold is set to 10% or 20% of the genome (of a presumably wild individual) thought to correspond to cultivated ancestry (Table B1). This use of model-based ancestry inference for identifying admixed individuals is overrepresented in microsatellite studies, which is perhaps not surprising since competing phylogenetic methods are unavailable for this type of marker. However, even in SNP studies, which are ideally suited to the analysis of site patterns and gene tree topologies through phylogenomic methods (see below), ancestry inference often supports the use of explicit statistical tests involving f- or D-statistics, and is used in combination to arrive at a global picture of population structure and gene flow.

Table 1. Non-exhaustive overview of analytical methods for detecting introgression, and their theoretical scope of application. Not all methods we mention in text were cited by the articles we surveyed; for those that were, we report the number of articles surveyed that cite the method as well as one example of its application.

Method	Type	Reference	Presence of gene flow	Population extent (admixed individuals)	Genomic extent (introgressed regions)	Direction of gene flow	Timing of gene flow	Number of gene flow events	Migration rate	Citing articles surveyed	Example of citing article
STRUCTURE	global ancestry	Pritchard et al. (2000)		x						24	Lima bean (Heredia-Pech et al., 2022)
ADMIXTURE	global ancestry	Alexander et al. (2009)		x						2	Carrot (Hernández et al., 2023)
fastSTRUCTURE	global ancestry	Raj et al. (2014)		x						1	Rice (Qiu et al., 2020)
InStruct	global ancestry	Gao et al. (2007)		x						1	Wheat (Pajkovic et al., 2014)
NGSadmix	global ancestry	Skotte et al. (2013)		x						1	Spearmint (Olofsson et al., 2024)
sNMF	global ancestry	Frichot et al. (2014)		x						1	Olive (Zunino et al., 2024)
DAPC	population clustering	Jombart et al. (2010)		x						0	n/a
Fuzzy c-means clustering	population clustering	Bezdek et al. (1981)		x						2	Wheat (Arrigo et al., 2011)
RFMix	local ancestry	Maples et al. (2013)			x					0	n/a
ELAI	local ancestry	Guan et al. (2014)			x					1	Maize (Le Corre et al., 2020)
HAPMIX	local ancestry	Price et al. (2009)			x					1	Maize (Hufford et al., 2013)
Loter	local ancestry	Dias-Alves et al. (2018)			x					1	Rice (Wedger et al., 2022)
D	site patterns	Green et al. (2010)	x							15	Rye (Rabanus-Wallace et al., 2021)
f3	site patterns	Peter (2016)	x							1	Maize (Hufford et al., 2013)
f4	site patterns	Peter (2016)			x					5	Rapeseed (Wang et al., 2023)
fd	site patterns	Martin et al. (2015)			x					3	Sunflower (Baute et al., 2016)
fhom	site patterns	Martin et al. (2015)			x					1	Wild strawberry (Feng et al., 2023)

HyDe	site patterns	Blischak et al. (2018)			x					1	Strawberry (Fan et al., 2024)
Df	site patterns	Pfeifer & Kapan (2019)			x					0	n/a
Dp	site patterns	Hamlin et al. (2020)			x					0	n/a
Dfs	site patterns	Martin & Amos (2020)				x				0	n/a
Dfoil	site patterns	Pease & Hahn (2015)				x				3	Rice (Choi et al., 2017)
Delta	gene trees	Huson et al. (2005)	x							0	n/a
D3	gene trees	Hahn & Hibbins (2019)	x							0	n/a
QuIBL	gene trees	Edelman et al. (2019)			x		x			1	Barley (Jin et al., 2024)
D1	gene trees	Hibbins & Hahn (2019)					x			0	n/a
D2	gene trees	Hibbins & Hahn (2019)				x				0	n/a
DIP	gene trees	Forsythe et al. (2020)				x				0	n/a
TreeMix	graphs & networks	Pickrell & Pritchard (2012)				x	x	x *		7	Helmstetter et al. (2020)
DIYABC	graphs & networks	Cornuet et al. (2014)				x	x	x *		1	Eggplant (Page et al., 2019)
PhyloNet	graphs & networks	Wen et al. (2018)				x	x	x *		2	Cabbage (Saban et al., 2023)
G-PhoCS	graphs & networks	Gronau et al. (2011)				x	x	x	x	1	Rice (Choi et al., 2017)
BayesAss	demographic models	Wilson & Rannala (2003)				x			x	1	Sorghum (Mutegi et al., 2012)
MIGRATE-N	demographic models	Beerli & Felsenstein (2001)				x			x	3	Almond (Delplancke et al., 2012)
F_{ST}	summary statistics	Weir & Cockerham (1984)			x				x	1	Sorghum (Adugna et al., 2017)
d_{yx}	summary statistics	Nei & Li (1979)			x				x	0	n/a
HIest	hybrid identification	Fitzpatrick (2012; 2013)		x						1	Coffee (Verleysen e al., 2024)
introgress	hybrid identification	Gompert & Buerkle (2010)		x						1	Apple (Feurtey et al., 2017)
NewHybrids	hybrid identification	Anderson (2008)		x						2	Lettuce (Uwimana et al., 2012)
parentage analysis, e.g. CERVUS	hybrid identification	Kalinowski et al. (2007)		x						2	Macadamia (O'Connor et al., 2015)
diagnostic alleles	hybrid identification	n/a		x						4	Yam (Scarcelli et al., 2017)

* These methods also infer the weight of the inferred gene flow events, as admixture proportions of the recipient population, as opposed to G-PhoCS, which infers a continuous migration rate.

Other methods related to this type of ancestry inference, but free from model-based biological assumptions, are based on sparse non-negative matrix factorisation (sNMF) (Frichot et al., 2014), discriminant analysis of principal components (DAPC) (Jombart et al., 2010), or fuzzy *c*-means clustering, a variation on *k*-means clustering (Bezdek, 1981). While these and the aforementioned methods all estimate global ancestry proportions or population membership probabilities for the entire genome, local ancestry inference (LAI) methods such as ELAI (Guan, 2014) and HAPMIX (Price et al., 2009), can infer the ancestry of specific genomic regions along chromosomes (Figure 3C). These methods were introduced more recently than methods to detect global admixture, and were less commonly used to screen for introgressed regions than other scans using *D*- and *f*-statistics (described below), among the studies we surveyed. A notable exception is the use of Loter (Dias-Alves et al., 2018) to demonstrate recent high admixture in weedy rice populations in the US (Wedger et al., 2022).

Phylogenetic methods

Species or populations have relationships that are often represented as a bifurcating tree or phylogeny. These phylogenies can be generated from sequencing data, assuming an evolutionary model of sequence evolution, to provide an estimate of the "true" species tree. A common challenge is that different parts of the genome can have distinct demographic histories or be subject to varying evolutionary forces, which means that the inferred tree topology depends on the region(s) of the genome used for inference. One common biological cause of different gene tree topologies is incomplete lineage sorting (ILS), whereby the alleles of genes coalesce at different rates than the overall species history, producing discordant gene trees. This is the result of a stochastic process in which alleles have a certain probability to be sorted into different lineages by chance, based on population size and divergence time. A second biological cause of discordant gene trees is introgression, whereby alleles are transmitted from one lineage to another after they have diverged, through one or multiple episodes of gene flow, which can produce patterns of reticulate evolution. Lastly, (convergent) adaptation can lead to homoplasy and discordant gene trees when selection results in sequence identity that is not caused by shared ancestry (whether through horizontal or vertical transmission). The latter differentiates identity by state (IBS) from identity by descent (IBD), which is limited to patterns of similarity that might conflict with the species tree topology, but are nonetheless inherited (either by ILS or introgression)

In addition to biological causes, differences in tree topologies can also be attributable to technical causes. These might lead to gene tree estimation errors as a result of insufficient data, evolutionary model misspecifications, sequencing errors or a mix of those (Cai et al., 2021). When such errors and the confounding effects of selection can be ruled out (e.g., by studying only neutral regions of the genome that experience little to no selection pressure), ILS can effectively serve as a baseline scenario for the inference of introgression, where the null hypothesis assumes that only ILS is responsible for gene tree discordance, and significant deviations from the expected site patterns under this null hypothesis are taken as evidence for introgression. This inference is based on the premise that ILS produces noise, but that this noise is more or less evenly distributed along the genome and produces symmetric patterns of discordant gene tree topologies among the lineages (Figure 3D). By contrast, asymmetric gene flow will result in some discordant topologies being more common than others - a fact that can be exploited to infer whether alleles have been shared in excess in one direction rather than another.

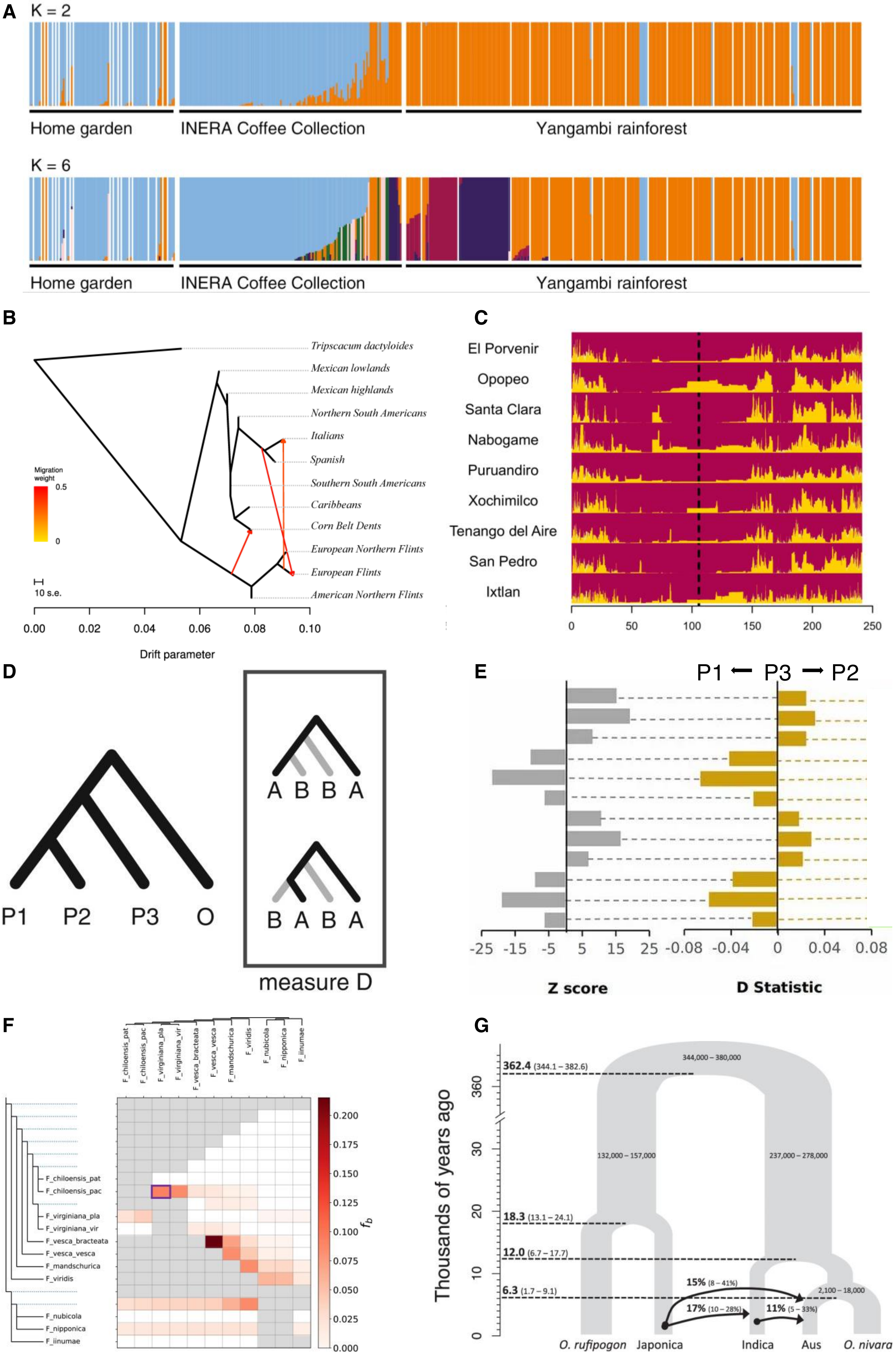

A
K = 2
Home garden
INERA Coffee Collection
Yangambi rainforest
K = 6
Home garden
INERA Coffee Collection
Yangambi rainforest
B
Tripscacum dactyloides
Mexican lowlands
Mexican highlands
Northern South Americans
Italians
Spanish
Southern South Americans
Caribbeans
Corn Belt Dents
European Northern Flints
European Flints
American Northern Flints
Migration weight
0.5
0
10 s.e.
0.00
0.02
0.04
0.06
0.08
0.10
Drift parameter
C
El Porvenir
Opopeo
Santa Clara
Nabogame
Puruandiro
Xochimilco
Tenango del Aire
San Pedro
Ixtlan
0
50
100
150
200
250
D
P1
P2
P3
O
A B B A
B A B A
measure D
E
P1 ⟵ P3 ⟶ P2
-25
-15
-5
5
15
25
-0.08
-0.04
0
0.04
0.08
Z score
D Statistic
F
F_chiloensis_pat
F_chiloensis_pac
F_virginiana_pla
F_virginiana_vir
F_vesca_bracteata
F_vesca_vesca
F_mandschurica
F_viridis
F_nubicola
F_nipponica
F_iinumae
0.200
0.175
0.150
0.125
0.100
0.075
0.050
0.025
0.000
f_b
G
Thousands of years ago
0
10
20
30
360
362.4 (344.1 – 382.6)
18.3 (13.1 – 24.1)
12.0 (6.7 – 17.7)
6.3 (1.7 – 9.1)
344,000 – 380,000
132,000 – 157,000
237,000 – 278,000
2,100 – 18,000
15% (8 – 41%)
17% (10 – 28%)
11% (5 – 33%)
O. rufipogon
Japonica
Indica
Aus
O. nivara

Figure 3 (previous page). Common methods used to detect and quantify gene flow. A) Global ancestry of *Coffea canephora* inferred with STRUCTURE with K=2 and K=6 ancestral populations (Verleysen et al., 2024). B) Population graph for maize populations inferred with TreeMix (Brandenburg et al., 2017). C) Local ancestry for *Zea mays* subsp. *mexicana* inferred with HAPMIX (Hufford et al., 2013). D) Principle of the *D*-statistic, also known as the ABBA-BABA test (Eaton & Ree, 2013). E) Measures of *D* for multiple population pairs of wild and domesticated *Zea mays* (Rojas-Barrera et al., 2019). F) *f*-branch statistic inferred for strawberry species with *Dsuite* (Fan & Whitaker, 2023). G) Demographic model inferred for wild and domesticated *Oryza sativa* with G-PhoCS (Choi et al., 2017).

Site patterns

In its simplest form, such a test examines many unlinked (i.e., freely recombining) biallelic sites in a four-lineage rooted tree (with three focal samples, representing distinct species or populations and an outgroup), where one of the three focal species shares an allele with the outgroup. Whereas BBAA represents the species or population tree, two alternative topologies are produced by ILS and/or gene flow, namely ABBA and BABA. The classic ABBA-BABA test calculates a test statistic *D* which is a measure of the imbalance between the ABBA and BABA topologies (Figure 3E): when it is positive, gene flow is needed to explain the excess of allele sharing between P2 and P3 (ABBA); when it is negative, gene flow is necessary to explain the excess allele sharing between P1 and P3 (BABA). Originally developed by Green et al. (2010), the *D*-statistic or Patterson's *D* is arguably the most commonly applied explicit test of introgression to date, with thousands of citations, and available in a range of different software implementations (Korneliussen et al., 2014; Malinsky et al., 2021; Patterson et al., 2012). The requirement of a large number of unlinked polymorphisms with only two alleles makes this test ideally suited for genome-wide SNPs.

While the idea is intuitive and *D* has proved to be remarkably robust to violations of underlying assumptions, it has some downsides: 1) although it can detect unidirectional gene flow, it cannot infer the direction; 2) it is not a reliable estimator of the genomic extent of gene flow; 3) it gives no information about the timing and mode (whether continuous or in pulses) of gene flow; and 4) it only considers three focal taxa. This does not have to be a problem if the only goal is to ascertain that gene flow happened - regardless of how much, when and in which direction. However, to satisfy user cases where the goal is to characterise these other dimensions of gene flow, the *D* statistic has been adapted into various alternative statistics that are able to characterise these different aspects. Many of these methods, as well as their strength and shortcomings, have recently been reviewed by Hibbins & Hahn (2022), but we will summarise some key differences here.

fD (Martin et al., 2015), *f4* (Peter, 2016) and *HyDe* (Blischak et al., 2018) estimate the extent of introgression by comparing the observed test statistic to a scenario where the entire genome would be introgressed. In contrast, *Df* (Pfeifer & Kapan, 2019) and *Dp* (Hamlin et al., 2020) are modifications of D that add the frequency of the non-introgressed topology to the denominator, thereby giving an estimate of the proportion of introgressed loci. D_{FS} (Martin & Amos, 2021) utilises the site-frequency spectrum (SFS) to infer the relative timing of introgression, calculating *D* for sites partitioned across different allele frequency bins; this approach is especially sensitive to recent introgression. Partitioned *D*-statistics (Eaton & Ree, 2013) and D_{FOIL} (Pease & Hahn, 2015) both infer the direction of gene flow by using additional information from a fifth taxon. Finally, *Dsuite* (Malinsky et al., 2021) was designed to enable computationally efficient calculations of statistics like *D* and *f4* across all combinations of tens to hundreds of lineages, directly from the variant call format (VCF) file (Figure 3F). This programme also includes the *f-branch* statistic, which can infer gene flow on internal branches of a phylogeny through the assessment of correlated *f4* ratios across multiple lineages (Malinsky et al., 2018). Nearly three-quarters of the SNP studies we surveyed employed an implementation of *D*- and/or *f*-statistics or one of their variations (Table B1).

Trees and networks

The *D* statistic ultimately only relies on biallelic sites without inferring gene trees. However, multiple substitutions in the DNA can cause patterns of sequence similarity that are not caused by descent from a common ancestor and hence lead to spurious site patterns, a phenomenon referred to as homoplasy. This problem increases with longer divergence times between lineages, but can be minimised by gene tree estimation methods that assume a suitable substitution model and can to some extent account and correct for multiple substitutions. Homoplasy is less likely to affect population or species pairs that recently diverged, rendering the correction of multiple substitutions a matter of less urgency and importance in the case of crops and closely related wild species, such as the wild progenitor. However, the recent divergence time in this case leads to other problems, as it increases incomplete lineage sorting, and the sequence similarity (caused by shared ancestry) makes it more difficult to detect "foreign" DNA.

When the genes used to infer trees freely recombine, and there is no recombination within them (the same assumptions that are made for biallelic SNPs), the resulting gene tree topologies and pairwise distances between lineages can be analysed in the same way as site patterns, without necessarily requiring an outgroup (Hahn & Hibbins, 2019). When branch lengths, which are indicative of coalescent times when measured in coalescent units, are also taken into account, even more information can be gleaned from the comparison of gene trees: the variation in coalescent times of different genes relative to the divergence time between lineages can be used to reveal the timing and extent of introgression; to infer the direction of gene flow (Forsythe et al., 2020; Hibbins & Hahn, 2019); to distinguish introgression from ancestral population structure, e.g. with QuIBL (Edelman et al., 2019); and even to detect introgression between sister taxa (P1 and P2): all things that site patterns alone, that are time agnostic, are not able to do. Despite these advantages, only one of the SNP studies in our survey using (variations of) the *D*-statistic used gene tree topologies or branch lengths to test for introgression (Jin et al., 2024). This could be because too little evolutionary time has passed since divergence for branch lengths to be meaningfully inferred or add much information; because homoplasy in these crop-to-wild systems is less of a concern; or simply because researchers have a preference for statistics based on site patterns due to their ease of use and familiarity.

The detection of discordant site patterns and tree topologies still rely on a backbone phylogeny or "true" tree relative to which the discordant topologies are identified. The most computationally advanced methods employ a (pseudo or full) likelihood or Bayesian approach to infer a phylogenetic network, jointly estimating evolutionary relationships, gene flow and a variety of other parameters in a single process - thereby bypassing the two-step process of first estimating a bifurcating tree and subsequently detecting hybridisation within a narrow set of constraints (Figure 3G). These methods are powerful because they can compare multiple scenarios and simultaneously identify the lineages that have introgressed as well as the extent, direction and timing of introgression, but they vary in the assumptions they make about the mode of introgression: whereas methods based on the isolation-with-migration (IM) model assume constant gene flow, other methods based on the multi-species network coalescent

(MSNC) assume discrete pulses. The only paper surveyed estimating phylogenetic networks with MSNC-based methods (Saban et al., 2023), namely PhyloNet (Wen et al., 2018), detected multiple reticulations in the history of wild *Brassica oleracea*. The only paper surveyed estimating continuous gene flow with an IM-based model, namely G-PhoCS (Gronau et al., 2011), studied the genetic contributions of domesticated *Oryza sativa* subsp. *japonica* to the wild progenitor of subsp. *indica*, suggesting that the two subspecies most likely share a single domestication event despite having multiple origins (Choi et al., 2017). However, other studies (Delplancke et al., 2012; Ha et al., 2021; Heredia-Pech et al., 2022) have made use of different Bayesian methods to estimate continuous migration rates, such as MIGRATE-N (Beerli & Felsenstein, 2001) and BayesAss (Wilson & Rannala, 2003). These tools can handle less complexity and make more assumptions than the IM model, but can be used with various types of markers in addition to SNPs.

3.3 Methodological advances and trade-offs

Despite their advantages, full-likelihood and Bayesian methods for detecting introgression are burdened by their high computational demands and limited scalability. Nonetheless, many advances are currently being made in the realm of Bayesian statistics regarding the inference of phylogenetic networks (Degnan, 2018; Elworth et al., 2019; Jiao et al., 2021). Their limited application therefore probably reflects a trade-off that is characterised by increases model complexity at the expense of data volume: while less sophisticated methods can be routinely applied to millions of markers and thousands of samples, the most complex model-based inferences can only be run for a handful of individuals using a subset of loci. Which method(s) should be used ultimately depends on the goals and expertise of the end user. This is in part determined by the complexity of the scenarios considered and whether gene flow is considered on a global scale or for specific parts of the genome. The use of a small number of loci may be suitable for detecting admixture, but unreliable for detecting its genomic extent. When using genome-wide data, reduced representation sequencing techniques can be an attractive option because of their relatively low cost, hence enabling the screening of larger numbers of individuals. This technique will usually ensure that the sequenced loci are unlinked, hence satisfying one of the common assumptions of introgression tests. However, the low density of SNPs can be a pitfall when inferring introgressed regions: smaller introgression tracts that fall between the sequenced loci might not be detected. The chosen study design is therefore the result of a balancing act between multiple considerations such as desired sample size (restricting the sequencing budget available per individual), desired genomic coverage (limiting the statistical power to detect introgression), and desired analytical technique (posing constraints on data type and limiting the amount of data that can be easily handled).

Finally, in an opposite trend from ever more complex models capturing different evolutionary processes, detection of all kinds of biological patterns is increasingly facilitated by the application of deep learning algorithms, which can be used to classify population genetic data without necessarily requiring theoretical knowledge of the processes that created them. While Flagel et al. (2019) already remarked on the potential of convolutional neural networks (CNN) to detect, for example, selective sweeps along the genome, more recently they have also been used to detect signatures of (adaptive) introgression (Gower et al., 2021; Zhang et al., 2023). To our knowledge, no one has yet attempted to apply these deep learning methods to the detection of crop-to-wild introgression.

3.4 Challenges in detecting introgression

Even though all methods described above differ slightly in the way they detect and characterise introgression, there are some common challenges. Firstly, they all suffer to some degree from confounding factors. First and foremost, ancestral allele sharing caused by incomplete lineage sorting (ILS) is the most common cause of different genealogies along the genome, especially between lineages that diverged only recently (such as crops and their wild relatives). But while almost all methods described here explicitly take ILS into account and take it as baseline for detecting introgression, the majority cannot rule out the possibility that selection or other causes of homoplasy create some of the signals picked up by discordant gene tree topologies, changes in allele frequencies or summary statistics. The presence or importance of selection on certain parts of the genome in the face of strong or recurrent gene flow is perhaps negligible on a genome-wide scale, but in localised regions these processes can produce misleading results. Other model violations could stem from linkage between loci, recombination within loci, or from the erroneous estimation of gene tree topologies. Many methods also assume that there is no ancestral population structure, an assumption that is often violated and can cause spurious signals of introgression where it hasn't occurred. An example of the possible confounding effect of ancestral population structure is offered in a study on wheat wild relatives by Huynh et al. (2019), who detected it by using a more distant outgroup. Few other studies explicitly examine and test these underlying assumptions.

In addition, admixture is often assumed to occur between distinct lineages or previously isolated populations. But past gene flow can obscure signals of more recent gene flow, and the relative proportions in each direction are often difficult to disentangle. In addition, introgression is harder to detect between lineages that are genetically very similar (such as crops and their wild progenitor, especially in their centre of diversity), and when the alleles that are introgressed in the receiving population are not fixed in the donor population. To make matters worse, introgression from unsampled or "ghost" lineages can lead to wrong inferences if one is not aware of the missing lineage that contributed alleles to the focal populations (Tricou et al., 2022). Some of these problems can be mitigated by explicitly testing for asymmetric gene flow, and by carefully selecting the populations to study and comparing the results when using different combinations of lineages. Crucially, identifying and minimising these confounding factors requires that the demographic history and relationships between populations are known, and that a suitable outgroup is used. Misspecification of the species tree, incorrect demographic scenarios, or sampling artefacts can therefore all bias the results.

In addition to knowing when the results are either false or biased (due to model violations or confounding factors), it is equally challenging to reliably interpret them when they are true. Observed patterns often have multiple explanations and it can be difficult to choose which is one is more likely. Results are particularly liable to overinterpretation when they are generated by methods that are model-agnostic or that rely on particular model assumptions rather than inferring these directly from the data (e.g., continuous gene flow), and subsequently used as evidence for the assumed model. Even when results are discussed in a nuanced manner, the visualisation of introgression as edges on a bifurcating tree often hints at findings that have not necessarily been substantiated or tested (i.e., that gene flow happened in a discrete event, at a particular point along a branch, or between particular lineages). Caution should therefore be exercised when visualising or looking at the results of introgression detection methods, as much as when describing them.

4 Prevalence of crop-to-wild introgression

One of the objectives of this study was to determine how common crop-to-wild gene flow is in different crop species. Broadly, our survey of literature from the past 15 years (2010-2024) shows that claims of crop-to-wild introgression have been made in a wide variety of cultivated species, covering 25 species in 23 genera and 13 plant families globally. Taking these studies at face value suggests that introgression is rampant across the plant tree of life, as not a single article studied introgression and found no evidence for it. However, doing so would ignore the fact that firstly, our survey is neither a complete nor a representative sample of cultivated plant species, and secondly, that biases in funding, reporting and publication could lead to an overrepresentation of positive results compared to instances where gene flow does not happen or is not observed.

4.1 Representativeness of studies

The reported results are skewed in multiple ways. Firstly, they are skewed geographically, with most studies originating from the US, Europe and Asia, while other continents - particularly Africa - are underrepresented. While our search was restricted to articles written in English, which should theoretically cover most peer-reviewed international journals, this could introduce some bias in terms the countries represented in the data. The results are also skewed in terms of taxonomy, with a relatively small list of crops (notably cereals) dominating the surveyed articles, especially those that correspond to or are closely related to genetic model systems. The vast majority of studies were on crop-to-wild gene flows in grasses (Poaceae), followed by legumes (Fabaceae), fruit trees (Rosaceae) and composites (Asteraceae). This could reflect economic interests in the concerned crops, (e.g., cereals are some of the most important food staples and their domestication histories have been intensively studied), but also availability of genomic resources, which form an important entry barrier to studying novel crop systems. Crops with larger genomes, longer generation times, or of lower commercial interest, are more likely to be neglected in terms of genomic resource development and scientific study. This taxonomic and geographic bias makes it difficult to ascertain in which crops introgression to crop wild relatives is most likely to occur, when the majority of them either have not been the subject of scientific study or are underrepresented in the literature.

In addition, for the crop systems that have been studied and are well represented, it could be possible that the results are skewed towards reporting evidence of gene flow rather than finding no evidence or not reporting it. Negative results could either not reach publication, in which case they would not be searchable in the databases we used, or they might simply not be reported (in sufficient detail) in the published studies at the expense of other results, in which case the study would not be selected as having investigated gene flow and/or not pass our eligibility criteria. Hence, the publication process and our literature screening protocol inherently imply that positive or significant results of gene flow are more likely to be found and included in our survey than negative or insignificant results. This does not diminish the fact that, when positive evidence is reported and is credible, we can learn valuable insights from these results; however, it does mean that we cannot take absence of evidence as evidence of absence.

4.2 Comparisons between studies

For the evidence that we found, the large diversity of methods used in the surveyed studies means that it is difficult to make direct comparisons between them. Even when the same statistic is used (e.g., D), it has been shown that statistical methods across different marker types (e.g., genotype-by-sequencing or whole genome sequencing) do not yield comparable results, with reduced representation methods on average producing higher values of D (Dagilis et al., 2021). Hence, the inferred presence and strength of introgression is dependent not just on the analytical methods used, but also on the type of sequencing data generated. Studies using similar data and similar methods might be used to infer differences in observed gene flow between systems, but are scarce. The main exception to this is the numerous studies published on crop-to-wild gene flow in apples, using the same microsatellite primers and the same method for inferring global ancestry and threshold for assigning admixed/hybrid status. This allows for direct comparisons within the same study system and has revealed some interesting dynamics of crop-to-wild introgression in different parts of Eurasia (Box 2).

Even when similar statistics and sequencing data are used to infer introgression, the results of tests for gene flow heavily depend on the selected lineages (i.e., the choice of crop cultivars and wild species considered) and population level sampling of these lineages. Introgression might occur in a specific local context where the crop and its relative occur in sympatry, but not in others (Hernández et al., 2023; Hufford et al., 2013). In addition, recent introgression might be missed when using samples that were collected decades or even centuries ago, such as those stored in gene banks. This is evidenced by the small number of studies that used a sampling design with an explicit temporal dimension, and that found a (sharp) increase in admixture over the last decades (Olodo et al., 2020; Spear et al., 2023). Without an explicit geographic and temporal dimension in sampling of both the crop and wild relative, the effects of place and time on the detection of introgression (and hence how commonly or likely it occurs) will not be given due consideration in assessments of the likelihood of gene flow.

In the ideal case, for each crop, detailed studies of global occurrences of gene flow would result in a probability map detailing the risk of introgression in specific places for specific crop wild relative species. The closest equivalents to such a map in any crop system are patterns of gene flow between *Malus domestica* and multiple of its Eurasian relatives (Box 2), and global patterns of gene flow in the evolution of weedy rice (Presotto et al., 2024; Qiu et al., 2020; Wedger et al., 2022). These studies show that the intra- and inter-generic variation in occurrence and extent of introgression between crops and wild relatives is large and highly context-dependent, and could exceed the variation observed between different crop systems. When only a handful of studies of introgression are available for each crop, all marked by incomplete sampling and a restricted geographic and temporal scope, the differences observed between crop systems (even when using similar data and methods) will be heavily influenced by the chosen sampling design and location, and unlikely to be generally applicable to the entire species.

4.3 General observations

Despite not being geographically or taxonomically exhaustive, the studies we surveyed still give some interesting insights into the presence of introgression, its extent and direction, the number of gene flow events and their timing, or the continuous migration rate. Here, we summarise the findings of the studies surveyed along these axes, illustrating them with examples from the literature. The dimensions and levels of introgression quantified by these and all other studies surveyed are detailed in Table B1. Some studies dealt with the same species and hence collectively shed light on the prevalence of introgression from the crop into its wild relatives. For this reason, species-specific summaries are provided in the Supplementary Information. An overview of all species reviewed, whether they have (priority) crop wild relatives in the Netherlands or in Europe, and the dimensions of introgression for which evidence was found is given in Table 2.

Presence

The studies we surveyed all in one way or another explicitly attempted to detect or quantify introgression. The most common method used was by inferring global ancestry. Sometimes this was the only quantifiable result reported. This was the case for most studies on apple (Bina et al., 2022; X. Chen et al., 2023; Cornille et al., 2013, 2015; Denoirjean et al., 2021; Feurtey et al., 2020; Omasheva et al., 2017; Schnitzler et al., 2014), a study on pearl millet (Olodo et al., 2020), one on weedy rice (Shivrain et al., 2010), and one on sunflower (Spear et al., 2023). Since ancestry coefficients can be explained by many factors, on its own this type of analysis offers limited evidence of gene flow. Hence studies basing their claims only on this method cannot conclusively determine whether the admixture patterns they report are the result of gene flow, without additional (non-genetic) observations.

In many cases, supplementary analyses were done, for example by estimating the migration rate, e.g., in almond (Delplancke et al., 2012), sorghum (Adugna & Bekele, 2017; Mutegi et al., 2012), and wheat (Pajkovic et al., 2014); by quantifying the genomic extent, e.g., in carrot (Hernández et al., 2023) and lima bean (Heredia-Pech et al., 2022); or by identifying hybrid categories, e.g., in apple (Ruhsam et al., 2019), coffee (Verleysen et al., 2024) or lettuce (Uwimana et al., 2012). Whereas the migration rate is almost at odds with the assumptions of global ancestry inference, which models recent admixture instead of continuous gene flow, the latter two provide meaningful ways to fine-tune this picture: detecting the genomic extent by identifying specific regions that are introgressed, and hybrid classification by subdividing admixed individuals into different stages of backcrossing. In the few studies we surveyed that performed these complementary analyses, their results are usually in agreement with the ancestry coefficients inferred by STRUCTURE-like algorithms, lending credibility to their conclusions.

Interestingly, the classical way to test for introgression, namely the ABBA-BABA test (or *D*-statistic), was in many cases not the first analysis performed. Rather, it was performed as a follow-up to confirm other results - for example, to identify regions in the genome with outlying *D* values in soybean (Wang et al., 2019), or to confirm instances of gene flow between specific subpopulations in rice (Qiu et al., 2020). Only five studies exclusively relied on the *D*-statistic, and were hence not able to characterise the genomic extent of introgression or determine the timing or direction of gene flow. These studies (Mondon et al., 2018; Olofsson et al., 2024; Presotto et al., 2024; Rojas-Barrera et al., 2019; Schreiber et al., 2019) tended to apply the ABBA-BABA test to different trios of populations, and quantify gene flow by the number of trios or population pairs for which *D* was significantly positive/negative. This was sometimes done to compare alternative scenarios of introgression (Presotto et al., 2024), to determine whether an observed introgression signal could be explained by unsampled lineages (Rojas-Barrera et al., 2019), or to estimate the proportion of individuals in a population that were introgressed (Olofsson et al., 2024).

None of the studies we surveyed searched for signals of introgression and found no evidence for it, hinting that the detection of introgression could be a matter of "seek, and you shall find" - although it could also mean that introgression is mostly studied in systems where it is thought to occur. However, despite the wide-spread positive evidence, the studies vary in how much introgression was found, its symmetry, the number of gene flow events detected (and their age), and/or the strength of migration.

Extent

Rather than simple presence or absence of gene flow between lineages, many studies were concerned with quantifying *how much* gene flow had happened. This was quantified in one of two ways: as a percentage of the population (expressed as the number of individuals in the population that carried signs of introgression, usually above some threshold value - although such a threshold was not always specified), or as a percentage of the genome (either per individual, averaged per population, or expressed as a range).

Population

More than twenty studies used global ancestry results to quantify the extent of admixture in a population, making it the most popularly used method. Percentages ranged from 2% to almost 100%, depending on the study system. Smaller sample sizes make it statistically more likely to observe extreme values, hence these figures cannot be easily compared. In addition, different threshold were used for classifying wild individuals as 'admixed', ranging from >10% to >30% cultivated ancestry. Admixed individuals frequently comprised up to a quarter, a third or even half of the populations, in crop systems as diverse as eggplant (Page et al., 2019), lima bean (Heredia-Pech et al., 2022), and carrot (Hernández et al., 2023). Reasonably low values (<10%) were found in *Coffea canephora* (Verleysen et al., 2024), *Lactuca serriola* (Uwimana et al., 2012) and from wheat into *Aegilops geniculata* (Pajkovic et al., 2014); the highest (>50%) in weedy rice (Wedger et al., 2022) and in six-rowed wild barley (Guo et al., 2022). The latter turned out to be a stable hybrid lineage, with 100% of individuals "admixed".

Population extent was sometimes also quantified with diagnostic alleles (such as chloroplast haplotypes or other types of markers), parentage analysis or hybrid classification. The lowest proportion of admixed individuals using these methods was 2%, in a wild *Helianthus petiolaris* population in Argentina (Gutierrez et al., 2010), which was much lower than found for a weedy *H. annuus* population in France, where the percentage of admixed individuals ranged from 7-38% (Roumet et al., 2013) or even 100% (Muller et al., 2011). Where measured, the proportion of populations harbouring admixed individuals was (much) higher than the proportion of admixed individuals in a population; e.g., 18% of individuals versus 67% of populations of *Aegilops triuncialis* were found to be admixed with wheat in a study in Spain (Parisod et al., 2013).

Genome

Genomic extent of introgression was often calculated in addition to a global ancestry inference, to scan specific regions that carried signs of introgression. Relying on genome-wide markers, these studies were mostly based on large numbers of SNPs, either generated by genotype-by-sequencing or whole-genome sequencing. Two studies used local ancestry inference. In Mexico, genomic introgression from cultivated maize into *Zea mays* subsp. *mexicana*, was estimated to be 9.2-11.4% (Hufford et al., 2023). In Arkansas (United States), the genomic extent of introgression from cultivated rice into weedy rice was estimated to be 26-31% (Wedger et al., 2022). Nine other studies used derivatives of *D*- or f-statistics. For example, the genomic extent of introgression from cultivated into wild lima bean (*Phaseolus lunatus*) was estimated with *Df* to be 12-18% (Heredia-Pech et al., 2022). Studies using *f4* found 9.46-14.28% genomic introgression from *Brassica oleracea* into *Brassica cretica* (Saban et al., 2023), 15.2-24.3% from cultivated into wild carrots in the United States (Hernández et al., 2023), 12-42% from maize into teosinte in Europe (Le Corre et al., 2020), and 7.1-9.7% from cultivated sunflower into its relatives, *Helianthus argophyllus* and *H. bolanderi-exilis* (Baute et al., 2016). A study on soybean using a custom statistical test based on relative frequencies of 'identical by descent' haplotypes (rIBD), found 0.059-41% (for an average of 1.9%) of the wild *Glycine soja* genome to be introgressed from cultivated *Glycine max*.

Direction and weight

Population graph methods, such as TreeMix, that rely on tree estimations and improve tree model likelihood by adding migration events, can be used to estimate the number of migrations, their direction, and their relative contribution to the recipient population (also called migration "weight"). Since TreeMix results are often visualised as a tree with migration vectors, drawn in a colour that represents the weight with a colour gradient (Figure 3B), numerical values are not always reported. However, for the studies that do, migration weights tend to be on the high side: e.g., a migration weight of 0.425-0.485 from cultivated into wild olives in the western Mediterranean (Zunino et al., 2024), and a migration weight of 0.14-0.39 from cultivated maize into teosinte in Europe (Le Corre et al., 2020). Using the Approximate Bayesian Computation (ABC) method DIYABC, Page et al. (2019) inferred the genomic contribution of cultivated eggplant (*Solanum melongena*) to its feral relative (*Solanum insanum*) to be 31-40%. Using G-PhoCS, Choi et al. (2017) inferred three migration events in a phylogenetic network of the three cultivated rice subspecies (*japonica*, *indica*, and *aus*) and their wild ancestors, with migration weights of 11% (*indica* to *aus*), 15% (*japonica* to *aus*) and 17% (*japonica* to *indica*), respectively.

Timing

The timing of introgression was rarely reported in the studies we surveyed, even if they were implicitly estimated with some of the methods used. For example, population graphs and demographic models such as G-PhoCS and PhyloNet allow for the inference of timing of introgression, by placing gene flow events on specific branches of the species tree or network. While these methods are able to distinguish gene flow events involving ancestral rather than extant lineages, other methods do not inherently account for this and can lead to spurious signals when introgression to or from ancestral lineages leave traces in multiple of their descendants. A common way to circumnavigate this issue, is to check whether signals of introgression (such as values of *D* significantly deviating from zero) are shared between close relatives. This principle is implemented in the software Dsuite, which visualises the *f-branch* statistic between population pairs on a tree, hence providing an intuitive way to detect clusters of positive signals among closely related samples. This method was employed by Fan et al. (2024) on wild strawberries, who found an ancient introgression between sympatric *F. chiloensis* subsp. *pacifica* and *F. virginiana* subsp. *platypetala*. In some studies, the authors also manually compared *D*-values between populations pairs to rule out confounding signals. This was done by Rojas-Barrera et al. (2019), who analysed introgression between historical and modern populations of wild and cultivated maize.

The timing of introgression can theoretically also be inferred from the genomic extent and the size of introgression tracts, which are expected to be large in the early stages following hybridisation and become smaller over time, although this was rarely considered among the studies we surveyed. The comparison of uniparentally and bi-parentally inherited markers can be very insightful in this respect, because chloroplast or mitochondrial haplotypes inherited from a maternal crop parent will persist over many generations, even if the signal on the nuclear genome has all but disappeared. This principle allowed Scarcelli et al. (2017) to classify wild yams into 'old' and 'recent' hybrids, depending on whether hybridisation was inferred only from the chloroplast haplotypes, or from nuclear microsatellites (or both).

Some studies expressly aimed to quantify admixture along a temporal dimension, and did so by analysing time series of samples over multiple decades. Examples of this are a study of pearl millet, showing homogenisation of the wild population over time (Olodo et al., 2020); a study on maize that quantified the genetic contribution of recently introduced modern maize varieties to traditional landraces and multiple subspecies of teosinte (Rojas-Barrera et al., 2019); and a study on sunflower that assessed the adaptation of urban sunflower populations to contemporary environmental change (Spear et al., 2023). These studies consistently demonstrate that crop-to-wild gene flow has increased in recent years, with little to no cultivated ancestry in decades old pearl millet and sunflower populations, compared to significant levels now, and with stronger D values in modern teosinte populations than in historical ones. While time series offer great potential for tracking introgression in real-time, it is more attainable and relevant for annual species rather than perennial species, which have longer generation times and where introgression necessarily happens over longer time scales.

Migration rate

Lastly, some studies quantified continuous migration rate as the number of migrants into the population (possibly per resident) per generation. For example, Ha et al. (2021) used MIGRATE-N to estimate the historical migration rate from cultivated apple into *Malus sieversii* in Kazakhstan to be 0.03-0.06 migrants per generation; Mutegi et al. (2012) used BayesAss to estimate the recent immigration rate from cultivated into wild sorghum at 0.08 migrants per generation; and Pajkovic et al. (2014) used Approximate Bayesian Computation (ABC) to estimate that the migration rate from cultivated wheat into *Aegilops triuncialis* was $2.7 \times 10{-}4$ migrants per resident per generation. Other studies report the mutation-scaled migration rate, which are often orders of magnitude larger and can vary widely (Adugna & Bekele, 2017; Delplancke et al., 2012; Heredia-Pech et al., 2022). As these numbers refer to a process (migration), it is not straightforward to directly link them to an outcome (introgression), which is per definition a pattern that results from the interplay of multiple processes. Low levels of immigration do not necessarily imply little introgression, and modest migration could conceivably still lead to substantial levels of admixture in the population or along the genome, depending on the temporal or spatial scale considered (Pajkovic et al., 2014). Most studies that estimated migration rates therefore did so in conjunction with other analyses that quantified population patterns of admixture, usually in the form of a global ancestry analysis.

Table 2. Species reviewed and their vernacular names, followed by whether they have crop wild relatives in the Netherlands or in Europe (with the number of species in parentheses), and evidence of introgression found and scored in literature. Narrative descriptions of the evidence per species are provided in Annex B; the extracted data for each study is provided in Table B1.

Species	Vernacular name	CWR in the Netherlands	Priority CWR in Europe	Evidence of introgression extracted from literature
Brassica napus	Rapeseed	Yes (4)	Yes (19)	Genomic extent; number of events
Brassica oleracea	Cabbage	Yes (4)	Yes (19)	Genomic extent; number of events
Coffea canephora	Coffee (robusta)	No	No	Population extent
Corylus avellana	Hazelnut	Yes	Yes	Number of events
Daucus carota	Carrot	Yes (1)	Yes (3)	Population extent; genomic extent; number of events
Dioscorea rotundata	Yam	No	No	Population extent
Fragaria × ananassa	Strawberry	Yes (2)	Yes (5)	Number of events
Fragaria vesca	Wild strawberry	Yes (2)	Yes (5)	Number of events; genomic extent
Glycine max	Soy bean	No	No	Genomic extent
Helianthus annuus	Sunflower	No	Yes (7)	Population extent; genomic extent
Hordeum vulgare	Barley	Yes (3)	Yes (5)	Population extent; number of events
Lactuca sativa	Lettuce	Yes (2)	Yes (11)	Population extent
Macadamia integrifolia	Macadamia	No	No	Population extent
Malus domestica	Apple	Yes (1)	Yes (3)	Population extent; migration rate
Mentha spicata	Spearmint	Yes (5)	Yes (1)	Population extent
Olea europaea	Olive	No	Yes (1)	Number of events; direction and weight
Oryza sativa	Rice	No	Yes (1)	Genomic extent; number of events; direction and weight
Pennisetum glaucum	Pearl millet	No	No	Genomic extent
Phaseolus lunatus	Lima bean	No	No	Population extent; genomic extent; migration rate
Prunus dulcis	Almond	Yes (3)	Yes (25)	Migration rate
Secale cereale	Rye	No	Yes (4)	Number of events
Solanum melongena	Eggplant	Yes (2)	Yes (5)	Population extent; number of events; direction and weight
Sorghum bicolor	Sorghum	No	Yes (2)	Genomic extent; migration rate
Triticum aestivum	Wheat	Yes (3)	Yes (26)	Genomic extent; migration rate
Zea mays	Maize	No	No (but see Box 2)	Genomic extent; number of events; direction and weight

4.4 What we don't see: filling in the blanks

What the preceding sections show, is that introgression between crops and close or distant wild relatives is incredibly common during crop domestication and dispersal. They also show that introgression does not necessarily involve only the wild progenitor, but also more distantly related wild species that can even belong to other genera altogether, such as in the case of wheat and its relatives (Arrigo et al., 2011; Pajkovic et al., 2014; Parisod et al., 2013). About a third of the studies we surveyed (also) considered introgression with relatives other than the wild progenitor, and about 20% of the studies explicitly considered multiple wild relatives. These samples are not sufficient to judge whether introgression happens any less frequently with more distant CWR, even though this is certainly expected due to increased divergence times, which tends to lead to stronger reproductive isolation (as explained in more detail in the next chapter). But while it is biologically expected that gene flow is more likely between species or populations that are less diverged, the higher degree of similar genomic similarity can simultaneously make it harder to detect. Many instances of crop-to-wild introgression - especially those involving the wild progenitor - might therefore go unnoticed, simply because they share a recent common ancestor and, hence, much of their genetic background. As an example, crop-to-wild gene flow between cultivated and wild olive in the western Mediterranean could be detected relatively easily because wild olives in this area have an ancestry that is distinct from the eastern Mediterranean, where cultivated olive was likely domesticated (Zunino et al., 2024)

While it remains difficult to make claims about the prevalence of introgression in crop systems that have not been studied or were not included in this survey, we can still draw some inferences based on basic principles of reproductive isolation, which are outlined in the section below. For example, many perennial crops such as fruit trees are known to be outcrossing and to have relied heavily on wild-to-crop gene flow during domestication (Gaut et al., 2015). It is therefore perhaps not surprising that gene flow in the other direction is also reported, and if studies on common fruit and nut trees are anything to go by (Cornille et al., 2014; Helmstetter et al., 2020; O'Connor et al., 2015; Zunino et al., 2024), we can expect to see a lot more gene flow between other perennial crops and their wild relatives.

Other crops are themselves the product of hybridisation, e.g. spearmint and strawberry (Fan & Whitaker, 2024; Olofsson et al., 2024), and they or their wild progenitors could potentially serve as bridges for genetic exchange to other wild relatives. Indeed, the fact that crops arose through hybridisation is itself a sign of the propensity of the wild progenitors (and potentially relatives) to hybridise and generate new lineages, which might mean that crop-to-wild gene flow is both expected and common in these cases (Fan & Whitaker, 2024; Olofsson et al., 2024). Surprisingly, the ploidy "barrier" (long thought to prevent gene flow between organisms of different ploidy levels) that sometimes arises when hybridisation results in allopolyploid offspring, does not appear to be wholly impermeable in some of these hybrid crop lineages, as gene flow has been observed to cross this barrier in multiple instances (Arrigo et al., 2011; Olofsson et al., 2024). This is consistent with emerging evidence from non-domesticated plant systems (Bartolić et al., 2024)

Even in clades where hybridisation and introgression are common, their prevalence and outcome may vary, depending on characteristics of the donor and recipient populations. In this respect, there are several particularities that collectively distinguish the scenario of a crop donor and wild recipient from other instances of gene flow (although some of them may be shared). These include an often low divergence time, especially between most crops and their wild progenitors; a small effective population size, but frequently high densities of the crop in cultivated areas; a relatively high genetic burden or load associated with the cost of domestication; possible ploidy changes, which are hypothesised to frequently play a role in domestication; and the presence of traits that are strongly selected for in an agricultural context, but may be maladaptive in the wild. General insights regarding these particularities, as well as specific cases where these particularities are demonstrated, are detailed in the section below.

5 Factors influencing crop-to-wild introgression

Much of the literature on crop-to-wild gene flow reviewed elsewhere (Ellstrand, 2003; Ellstrand et al., 2013) has focused on determining whether crosses between crops and their wild relatives are even possible, and quantifying the viability and fitness of the resulting hybrid generations. These two topics broadly correspond to the study of pre-zygotic mechanisms of reproductive isolation (that prevent fertilisation), and post-zygotic mechanisms of reproductive isolation (that limit the survival and reproductive success of the resulting offspring). Our systematic search recovered several dozens of such studies, which often failed our eligibility criteria because they did not genetically analyse the fate of hybrids or introgressed regions under natural conditions. While many of these studies are experimental in nature and do not necessarily analyse molecular data, they are nonetheless a potentially useful resource by providing important baseline information on the level of reproductive isolation between species, which can explain some of the patterns of detected introgression that we see in the literature. Since we did not survey these studies here, we are unable to summarise this baseline information for individual species. However, below we outline the general mechanisms of pre-zygotic and post-zygotic barriers that can limit gene flow in plants.

5.1 Pre-zygotic barriers

Mechanisms of reproductive isolation can occur either before fertilisation (pre-zygotic) or after fertilisation (post-zygotic). Pre-zygotic barriers often play a larger role in reproductive isolation than post-zygotic barriers (Christie et al., 2022; Lowry et al., 2008). They also tend to be more pronounced in annuals than perennials, where outcrossing is the dominant mating system (Gaut et al., 2015). Pre-zygotic barriers can be related to geography, phenology, ecology, and reproductive biology.

Geography

In geographic terms, plants need to have overlapping distributions, or be within pollination distance from one another. In Mexico, Rojas-Barrera et al. (2019) observed gene flow between sympatric populations of maize and teosinte, but not in allopatric populations. In sympatric plant populations, the probability of gene flow between plants is a function of distance and dispersal capacity. This function follows leptokurtic curve, with a peak near the source and a long, fat tail (Ellstrand, 2014). This implies that the probability of gene flow declines with increasing distance, with the exact distance depending on the dispersal capacity of the species in question.

Consistent with this expectation, the incidence of admixed crab-apples was higher in populations in closer proximity to apple orchards (Feurtey et al., 2020). Similarly, Verleysen et al. (2024) found admixed *Coffea canephora* in the Congolese rainforest, but exclusively close to home gardens where the same species is also cultivated. Despite the lower probability of gene flow over longer distances, the sum of these small probabilities over large numbers of individuals can still be non-negligible. This is shown in the case of wheat introgression, where a low migration rate still led to noticeable levels of admixture in *Aegilops* populations, presumably due to high pollen pressure (Pajkovic et al., 2014). Density dependence is also illustrated in the case of apple orchards, where Cornille et al. (2015) found that cultivation density had a positive correlation with the rate of introgression in surrounding wild apple populations.

Some species have a natural capacity for long-distance dispersal, generally increasing with plant height and with faster life history strategies (Beckman et al., 2018; Thomson et al., 2018). Carried by animal vectors, wind or water, pollen and seed can sometimes travel many miles, allowing gene flow between what are sometimes considered to be geographically isolated populations (Buschbom et al., 2011; Gerber et al., 2014). This capacity for long-range dispersal can make it harder to detect possible source and sink populations in introgression studies, since the area to cover increases exponentially with a wider radius from the source.

Anthropogenic environmental pressures such as climate change and human-mediated dispersal can also change the distribution of species, leading to range shifts as well as introductions of species in new areas and even new continents. This could affect the co-occurrence of crops and their wild relatives. Due to human-mediated introductions, crop-to-wild and crop-to-weed gene flow has now been detected for a number of species outside their native range, including rice in North and South America (Wedger et al., 2022; Presotto et al., 2024), and teosinte and sunflower in Europe (Le Corre et al., 2020; Roumet et al.; 2013).

Phenology

In addition to physical proximity, fertilisation also requires overlap in flowering times and, in the case of an animal vector, shared pollinators - although in some species, manual pollination has replaced natural crop pollination, and pollen can even be stored and imported, theoretically eliminating the need for male plants (Chao & Krueger, 2007). Flowering time divergence has been confirmed to play a role in (partial) reproductive isolation between cultivated and weedy sunflowers in Europe (Roumet et al., 2013). An important crop trait determining growth and yield, flowering time has often been under strong selection during crop domestication and adaptation (Lin et al., 2021). As climate change is pushing back the flowering times of many wild plants (Geissler et al., 2023), this may reduce the phenological isolation between crops that have been selected for early flowering and their wild relatives.

Ecology

Crop-to-wild gene flow can be mediated either by pollen dispersal or by seed dispersal. Both of these can be non-random, leading to structured migration (Ellstrand, 2014). Pollen-mediated gene flow happens when wild plants are pollinated with pollen from cultivated plants. This can be highly asymmetric, depending on the pollen production of the plants and the density of the populations. In some perennial fruit-bearing crops, such as date palm, the density of males in the population is deliberately kept low, lowering pollen production and potentially minimising crop-to-wild gene flow (Chao & Krueger, 2007). Vice versa, manual pollination of plants allows control over the reproduction of the plant, preventing wild-to-crop gene flow. For animal-pollinated species, gene flow depends on shared pollinators - a condition that is for example satisfied in the case of cultivated and wild sunflowers, but that could be disrupted when crops are introduced outside of their native range where these pollinators are not present, as is the case for vanilla (Karremans, 2024). Pollinator populations are also increasingly under pressure from human activities; this may limit gene flow in areas with high levels of disturbance (Cornille et al., 2015).

Gene flow can also occur through seed dispersal. Cultivated seeds can be recruited by natural populations and introgress locally, as is observed in the case of macadamia (O’Connor et al., 2015). Another possibility is that cultivated plants are pollinated with wild pollen, and that the resulting admixed individuals subsequently come into contact with wild populations. This process is thought to play a role in the emergence of weedy sunflowers in sunflower fields in Spain and France (Muller et al., 2011), but also crop-wild hybrids of yam in West Africa (Scarcelli et al., 2017). Whether the crop parent was maternal or paternal can be established with the use of uniparentally inherited crop-specific markers, such as mitochondria and chloroplasts, provided these are sufficiently divergent between cultivated and wild populations. Few of the studies we surveyed have employed this approach, but those that did shed light on the biological mechanism of gene flow and are sometimes able to uncover more ancient hybridisation events that are no longer detectable with nuclear markers (Scarcelli et al., 2017).

Agricultural practices may influence the likelihood of seed-mediated versus pollen-mediated gene flow. In shifting cultivation, for example, farmers tend to abandon fields after several years of cultivation, providing a route for the remaining seeds of cultivated plants to come into contact with wild populations. Other human activities such as transportation over sometimes large distances, can facilitate accidental introductions of (genetically engineered) crop seeds in areas where they are not previously cultivated or authorised (Rostoks et al., 2019).

Reproductive biology

Another factor that determines the prevalence of gene flow is to what degree a species is self-pollinating. Outcrossing species or species with a mixed mating system are more prone to hybridisation with other populations or species than those that are obligate selfers. Mating systems as a trait is subject to evolution and self-incompatibility is controlled by a single locus several crop species, including apple (Cornille et al., 2014) and *Brassica* (Murase et al., 2020). Introgression of this allele into wild population may further increase the risk of crop-to-wild gene flow. Crop propagation methods also play a role, with sexual reproduction less common in crops that are vegetatively propagated. In vegetatively propagated crops, however, sexual reproduction can still occur, whether voluntarily or accidentally. This is demonstrated by the extensive crop-to-wild gene flow that has been observed in yams (Scarcelli et al., 2017). Even when outcrossing is rare and the immigration rate is theoretically low, introgression can still be common. Rapid introgression in a selfing plant has, for example, been demonstrated in the case of the wheat wild relative *Aegilops triuncialis*, where Pajkovic et al. (2014) calculated that the observed patterns of admixture (10% of individuals per population) could arise in less than twenty years.

5.2 Post-zygotic barriers

Post-zygotic mechanisms of reproductive isolation play a role after fertilisation, and manifest themselves either in the viability of hybrid offspring, their fertility and through hybrids fitness effects. Hybrid fitness effects can be positive (leading to hybrid vigour or heterosis), negative (leading to outbreeding depression or, in the longer term, hybrid breakdown), or a mix of both; but in the context of reproductive isolation there is usually a negative effect. Reduced hybrid fitness can arise through multiple mechanisms, sometimes involving complex genetic interactions between multiple alleles, either through hybrid incompatibilities such as Dobzhansky–Muller incompatibilities, hybrid load (caused by an accumulation of unlinked deleterious alleles, much like the genetic load associated with domestication), or segregation load, involving a de-coupling of co-adaptive alleles (Moran et al., 2021). None of the studies we surveyed looked deep into the genetic mechanisms of hybrid fitness effects. However, there are some organism-specific factors that influence viability, fertility and fitness of hybrid offspring.

Divergence

The relationship between genomic divergence and hybridisation tends to be noisy, since many biological factors affect the incidence of hybridisation. But despite the uneven effect of genomic divergence on reproductive isolation across taxa, higher divergence between taxa is generally associated with a lower incidence of hybridisation. This is partially explained by the fact that Dobzhansky-Muller incompatibilities are expected to increase with genetic distance; a phenomenon that has been termed the snowball effect. However, this prediction assumes that these incompatibilities are randomly distributed throughout the genome, and hence might not hold if they are concentrated in hotspots instead (Moran et al., 2021). Hybridisation at higher levels of divergence is often thought to be more common in plants than in animals. However, a recent study debunks this notion by demonstrating the plants experience less introgression at the same divergence levels, showing that plants are less promiscuous than once thought (Monnet et al., 2023). Nonetheless, the studies we surveyed exposed multiple instances of interspecific gene flow, e.g. from wheat (Arrigo et al., 2011), sunflower (Baute et al., 2016; Gutierrez et al., 2010), apple (Cornille et al., 2013) and rye (Rabanus-Wallace et al., 2021), where different plant species did not reach full reproductive isolation.

Genome structure

Genetic incompatibilities can also arise through differences in genome structure and function, which may evolve over time. Genomes of different ploidy or size, with a different gene content, or with structural rearrangements may be incompatible. Especially differences in ploidy and chromosome number are thought to pose a barrier to reproduction, and has been mentioned as a possible mitigation strategy for minimising crop-to-wild gene flow (Kwit et al., 2011). As some of the studies we surveyed show, this barrier can be highly porous and is no guarantee that crop-to-wild gene flow will not occur between species of different ploidy levels (Arrigo et al., 2011; Olofsson et al., 2024). Genomes with fewer and larger chromosomes exhibit more linkage between alleles, which leads to higher predicted levels of purging following introgression (Veller et al., 2023). Species with large genomes or chromosomes may therefore be more resistant to introgression.

Selection and recombination

The fate of introgressed regions or "introgression tracts" in a genome is determined by a complex interplay of selection, recombination and drift. When the recombination rate is low, negative selection quickly purges introgression tracts due to the presence of linked, deleterious alleles. But when deleterious alleles are absent or negative selection is weak, introgression tracts may persist, and in the case of positive selection even spread throughout a population. While single loci large effect sizes are theoretically rare in natural populations, crop improvement programmes have introduced some alleles into cultivars that could confer large fitness advantages on wild or weedy individuals. In multiple of the studies we surveyed, crop-to-wild gene flow was linked to the introgression of mutant herbicide resistance alleles that were originally introduced to control the very weeds that are now growing resistant to them (Le

Corre et al., 2020; Wedger et al., 2022), demonstrating adaptive introgression can occur rapidly when selection is strong.

When recombination is high, introgression tracts quickly break down and become smaller, unlinking potentially adaptive from deleterious alleles. This will preserve a stronger signal of introgression in the genome in the case of negative selection, and a weaker signal of introgression in the case of positive selection, than in regions where the recombination rate is high. As stated above, recombination rates are affected by genome size and function, and hence not all genomes may be able to purge deleterious alleles as easily as others. The relationship between recombination rate and introgression therefore depends on the strength and direction of selection: it is positive for deleterious alleles, and negative for beneficial alleles. A demonstration of this relationship is given in a study on diploid strawberries, in which the authors found that introgression was concentrated in regions of high recombination and low gene density, suggesting negative selection on introgressed alleles (Feng et al., 2023). The same relationship was observed in a study on *Brassica napus*, where introgression was also localised in regions of the genome with high recombination (T. Wang et al., 2023).

Demography

In addition to selection and recombination, the fate of introgression tracts depends on the effective population size[3], which mediates the effectiveness of selection through genetic drift. In areas of the genome with a higher recombination rate, the effective population size is larger and the impact of drift lower, making selection more effective. But the effective population size of the donor and recipient populations also matter. In smaller populations, the effects of genetic drift are known to increase the mutation load. Due to the severe bottleneck that many crops - and especially improved varieties - experienced during domestication and subsequent human-mediated selection, they are expected to contain an excess of deleterious alleles compared to their wild relatives (which tend to harbour more diversity and have higher effective population sizes). This would lead to overall negative fitness effects of crop-to-wild gene flow. If the effective population size of the wild recipient is large enough, the majority of introgressed deleterious crop alleles will most likely be removed from the wild population, especially in coding and conserved non-coding regions of the genome (Moran et al., 2021). However, vulnerable wild populations with a smaller population size may purge these alleles less effectively and therefore be more susceptible to the accumulation of (weakly) deleterious alleles due to introgression. This effect of demography is apparent in the case of almond, where comparable population sizes and levels of diversity lead to roughly symmetric gene flow between wild and cultivated almond (Delplancke et al., 2012). Similar observations have been made in other outcrossing perennials, such as hazelnut, where domestication has not caused a strong reduction in diversity (Helmstetter et al., 2020). In these systems, crop-to-wild gene flow may be relatively harmless. In contrast, in severely fragmented wild apple populations, levels of reported admixture are much higher than in other perennials; here, there is a real concern that crop-to-wild introgression might overwhelm what is left of the wild species (Cornille et al., 2014).

---

[3] This concept is not to be confused with absolute population size or "census size" related to the density or abundance of crops described earlier. Effective population size is a measure linked to the level of inbreeding in a population and can be low, even when the absolute number of individuals in a population is (very) high.

Environment

Lastly, the fitness effects of introgression are also mediated by the environment, which can influence the direction and strength of selection. Crop-to-wild introgression can therefore produce different outcomes in different settings. An example is given by the contrasting effects of crop-to-wild introgression from maize into teosinte in Mexico and in Europe. In Europe, where teosinte was only recently introduced, it has rapidly evolved to become a weed. This was aided by introgression of flowering time genes and herbicide resistance from improved cultivars, facilitating its range expansion and adaptation to the local environment (Le Corre et al., 2020). In Mexico, where teosinte is native, crop-wild introgression has occurred in both directions for millennia and both cultivated varieties and wild populations are locally adapted. Here, wild-to-crop introgression has helped maize to adapt to high-altitude environments, and crop-to-wild introgression is less commonly observed than the other way around (Hufford et al., 2013). In addition to differences in environment experienced by contemporaneous populations in distant localities, climate change can alter the possible fitness effects of crop alleles over time. It is conceivable that crop-to-wild gene flow can accelerate the adaptation of wild population to changing environmental conditions - a possibility that was explored in a resurrection experiment of urban sunflowers in the United States, which showed reduced phenotypic plasticity in flowering time that may have been the result of gene flow from cultivated sunflowers (Spear et al., 2023).

6 Consequences of crop-to-wild introgression

With the increasing realisation that gene flow during and after speciation is much more common than previously thought, scientific research has started to shift its focus away from whether introgression occurs, to what its effects are and how its dynamics are played out (Arnold & Kunte, 2017). The myriad of processes at play described above, which influence the likelihood and trajectory of introgression, makes it difficult to predict for any one system what its consequences will be. The diversity of evolutionary consequences of crop-to-wild introgression have been described in a special issue on the importance of gene flow in applied biology (Ellstrand & Rieseberg, 2016). The articles in this issue underscore the uncertainty of the impact of gene flow on hybrid fitness, which can either increase, decrease, or stay the same. Empirical and theoretical advances, however, increasingly shed light on the genomic dynamics of introgression following hybridisation. These show that hybridisation can be seen as a shock to the genome, after which it needs to stabilise and find a new equilibrium (Moran et al., 2021). Where this equilibrium lies depends on the size of the initial shock, different population parameters and on the environmental context.

Below, we classify the possible outcomes of crop-to-wild gene flow by their effects on the crop wild relative population. While not many studies in our survey explicitly addressed these consequences, where possible we will link them to examples from the literature. In brief, if the strength of gene flow is too strong for a crop wild relative to maintain its genetic integrity, it may suffer from genetic erosion and even go extinct (demographic swamping). It may also lead to the creation of a new, hybrid population, that could co-exist with the wild parents or outcompete and eventually replace them (genetic assimilation). In contrast, when gene flow leads to the introgression of beneficial alleles, crop wild relatives may acquire novel traits that enhance their fitness. These could either be of conservation value, by increasing the adaptive potential of natural populations (genetic rescue), or have detrimental effects by increasing invasiveness or weediness. Lastly, introgression can sometimes confer a set of domestication traits known as the domestication syndrome, and lead to the emergence of newly domesticated lineages. Three examples with contrasting outcomes of crop-to-wild introgression are highlighted in Box 2.

6.1 Genetic assimilation

Genetic assimilation or genetic "swamping" occurs when the recipient population is overtaken by individuals of admixed ancestry. In the most extreme case, a hybrid lineage can replace both its parents and lead to the extinction of the parental genomes, while giving rise to a new and stable lineage. While a scenario like this is most likely in the case of hybrid vigour or heterosis, i.e. when hybrids have a higher fitness than either of their parents, increased fitness is not required for genetic assimilation to take place, as long as there are no strongly deleterious effects and gene flow is strong enough (Hernández et al., 2023). Among our surveyed studies, we encountered extremely high proportions of admixed individuals (more than half of the population) in only four cases. Two of these described crop-to-weed gene flow in rice, with 53% of individuals classified as admixed in Latin American weedy rice (Qiu et al., 2020) and 73% classified as admixed in Arkansas (Wedger et al., 2022). In the latter, weedy rice picked up herbicide resistance from a local cultivar, demonstrating the adaptive significance of the introgressed alleles.

Box 2. Model systems for studying crop-to-wild introgression and case studies

Some of the most widely studied crops in the body of surveyed literature were apple (11 studies), sunflower (6 studies) and maize (3 studies). In addition to weedy and cultivated rice (6 studies), these crops are emerging as model systems for the study of intra- and inter-specific gene flow, benefiting from large research communities investigating the evolutionary dynamics of these crops and their wild relatives in different geographic regions and environmental settings, including Europe. The articles focusing on these crops are qualitatively different, using different data and approaches to shed light on the phenomenon of crop-to-wild introgression, and leading to different results and conclusions regarding its significance. Below we describe three case studies in more detail, that were selected for their contrasting methodologies, complementary strengths and divergent evolutionary outcomes.

Widespread introgression carries conservation implications for multiple wild apple species

In a seminal study by Cornille et al. (2013), crop-to-wild gene flow was investigated from domesticated apple into three of its wild relatives, including the European crab-apple (*Malus sylvestris*), a wild apple species from the Caucasus (*M. orientalis*), and the Central Asian progenitor (*M. sieversii*). As one of the earlier articles included in our survey, this study is characterised by a broad geographic coverage and a large sample size, including hundreds of individuals from dozens of sites per species. In total 1181 individuals were genotyped with 26 microsatellite markers to characterise admixture proportions with the Bayesian clustering algorithm STRUCTURE (Pritchard et al., 2000), taking care to assess population structure of the wild species prior to considering any domesticated individuals in order to determine a baseline number of ancestral populations K, before adding an additional population of pure domesticated individuals, that were confirmed not be admixed in previous analyses (Figure 5A).

Their results show that spurious ancestry coefficients can arise from setting the number of ancestral populations K too low, but that reliable admixture are obtained at baseline $K + 1$. This approach, while allowing some variation in the threshold used to identify admixed individuals, subsequently became the template for the crop-to-wild introgression studies on apple that followed. These studies exposed varying degrees of admixture in wild apple populations, ranging from as low as 3% to as high as 95%, and with different evolutionary outcomes. While crop ancestry in *M. sylvestris* has been associated with higher fitness (Feurtey et al., 2017) and increased aphid resistance (Denoirjean et al., 2021), introgression is widespread both in Europe (X. Chen et al., 2023) and Britain (Ruhsam et al., 2019) leading to concerns over reduced diversity. Regional pockets with low incidence of introgression have therefore been proposed to be of strategic importance for wild *Malus sylvestris* conservation (Schnitzler et al., 2014). Introgression rates in *M. orientalis* and *M. sieversii* were generally lower, but highly variable for the wild progenitor in Kazakhstan, where habitat loss severely fragmented the remaining populations and crop-to-wild gene flow is considered a risk to its genetic integrity (Omasheva et al., 2017). In another study, crop-to-wild gene flow was implicated in the spread of more virulent pathogens, posing an additional threat to the survival of wild populations (Feurtey et al., 2020). **Together, these studies show the conservation implications of crop-to-wild gene flow and the possible threat of introgression for genetic diversity of multiple crop wild relatives across the cultivation range.**

Maize introgression contributes to the emergence of a novel weed outside its natural range

The possibility of hybridization between cultivated and wild relatives is constrained by the distribution of wild relatives. But long-range dispersal, which is often human-mediated, has created opportunities for hybridization between wild and cultivated species outside their original range. Teosinte (a multi-species complex which can refer to either *Zea mays* subsp. *mexicana*, *Zea mays* subsp. *parviglumis*, or *Zea mays* subsp. *huehuetengensis*) are the closest relatives of cultivated maize, with an original distribution in Mexico and Guatemala. Two recent studies document the presence of these wild relatives in Europe in Spain and France (Le Corre et al., 2020; Trtikova et al., 2017). Using 24544 genome-wide SNPs of 1005 individuals, the authors established the genetic proximity of Spanish and French teosinte to the high-altitude teosinte *Zea mays* subsp. *mexicana*, based on principal component analysis (PCA), Bayesian inference of population structure (Pritchard et al., 2000), and analysis of gene flow with TreeMix (Pickrell & Pritchard, 2012).

In addition to establishing the descent of European from high altitude teosinte, these studies also provide evidence of hybridization with cultivated maize in both countries. Hybridization between French teosinte and cultivated maize is associated with the introgression of a local crop allele (of ZCN8 on chromosome 8) allowing flowering in temperate regions. Herbicide resistant maize was used to control the presence of teosinte in agricultural fields. Teosintes in France have subsequently acquired the herbicide resistance allele (of acetyl-CoA carboxylase 1 on chromosome 2) of European cultivated maize (Figure 5B). Analysing the exact regions where these introgressions occurred was enabled by inference of local ancestry with ELAI (Guan, 2014). **These results demonstrate the unique evolutionary trajectory of a crop wild relative outside of its native range, and how rapid adaptation to novel environments following range expansion can be fuelled by the introgression of locally adapted alleles as a source of evolutionary novelty.**

Rapid evolution of climate adaptation driven by crop-to-wild gene flow in urban sunflowers

Conservation of crop wild relatives (CWRs) is often seen as instrumental for crop improvement, harbouring a large genetic diversity and adaptive potential that could improve the resilience of crop varieties, particularly in the face of changing environmental conditions. Vice versa, the adaptive potential of crop diversity for wild populations is rarely explored. In one notable exception, cultivated sunflower diversity was implicated in ongoing climate adaptation of urban wild sunflowers (Spear et al., 2023). In a resurrection experiment performed on seed collections of *Helianthus annuus* from two populations in Minneapolis (United States), spanning several decades, Spear et al. (2023) directly compared the transcriptomes and fitness related traits of antecedent and modern individuals collected from the same sites, in order to assess their recent evolution.

Despite the relatively short time that passed between the oldest and youngest seeds (36 years), significant phenotypic changes were found, with a trend towards cultivated phenotypes and reduced phenotypic plasticity in flowering time. When exposed to different environmental watering and temperature regimes, modern populations were better adapted to current climate conditions than the antecedent population. This was coupled with a clear segregation of modern and antecedent populations in a principal component analysis (Figure 5C) and an increased proportion of cultivated ancestry in the genome, based on 3427 transcriptome-derived SNPs analysed with STRUCTURE (Pritchard et al., 2000). Although the authors did not identify introgressed regions, the combination of phenotypic selection on traits in the direction of cultivated varieties and increased cultivated ancestry in modern seed collections (which was nearly absent in the antecedent seed collection) points to the role of crop-to-wild gene flow in some of these adaptations, which appear to be driven by changing environmental conditions. **This study demonstrates the power of time series analyses to uncover rapid genetic and phenotypic changes, and highlights the potential role of crop-to-wild introgression in accelerating adaptive evolution of crop wild relatives in response to climate change.**

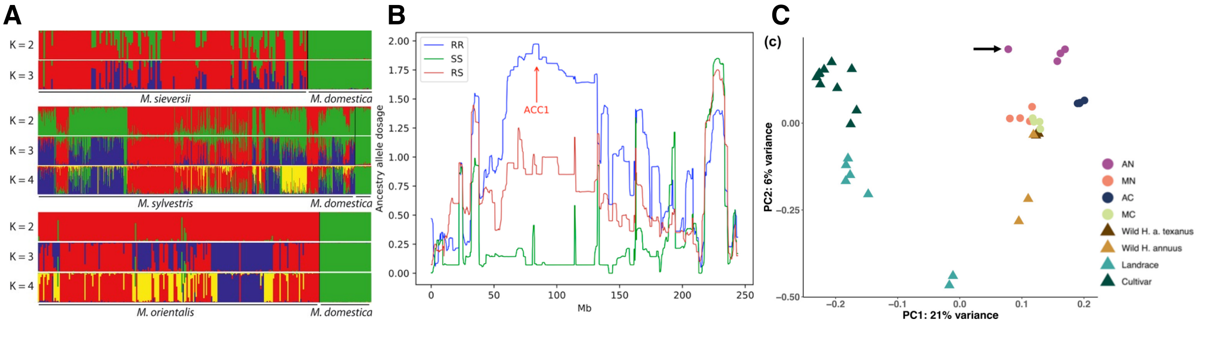


Figure 5. Detection of introgression in selected case studies. A) Cultivated ancestry in three wild species of apple (Cornille et al., 2013). Analysis of introgression of the *ACC1* mutant allele in chromosome 2 of teosinte (Le Corre et al., 2020)**.** PCA of modern and antecedent urban sunflower populations relative to wild and cultivated reference samples (Spear et al., 2023).

Guo et al. (2022) found that almost all sequenced individuals of six-rowed wild-growing barley (*Hordeum agriocrithon*), which has been hypothesised to be an ancestor of domesticated six-rowed barley, are in fact hybrids between cultivated six-rowed barley (*Hordeum vulgare*) and its two-rowed wild progenitor (*Hordeum vulgare* subsp. *spontaneum*); the small number that were not admixed were de-domesticated barley. The hybrid has co-existed with its parent species for a long time, and despite having multiple origins, which indicates spontaneous hybridisation across the parental range, there is no evidence that these six-rowed hybrids are actually replacing their two-rowed ancestors. In contrast, crop-to-wild gene flow was found to lead to varying proportions of admixed individuals in native populations of the wild progenitor of apple (*Malus sieversii*) in Kazakhstan, with one population reaching 95%; in this population, genetic swamping seems nearly complete, and might foreshadow the fate of other populations in the region if crop-to-wild gene flow is not managed, raising concerns over the survival of this species.

Another notorious example of genetic assimilation is the case of Californian wild radish. In this case, the hybridisation of the commonly cultivated *Raphanus sativus* and its wild relative *Raphanus raphanistrum*, both of which originated in Eurasia but were introduced to and became naturalised in the United States, led to the local extinction of both of its parents in the wild (Hegde et al., 2006). This result contrasts sharply to its native Europe where both radish species co-exist and occasionally hybridise without substituting the parent lineages. While the AFLP markers used to identify these American hybrids were not sufficient to pinpoint the regions of the genome involved, the authors hypothesised that their genetic success was the result of a novel combination of traits, encountered in neither of the parents. A similar observation has been made for wild carrots (*Daucus carota*) in two regions in the United States (California and Nantucket), where pure wild carrots no longer seem to exist as a result of crop-to-wild gene flow (Hernández et al., 2023). The distinct pattern of introgression left in the genome, detected by genome-wide SNPs, suggests that this replacement happened under neutral conditions. The authors speculate that the same introgressed alleles might be deleterious in other populations in the United States, where hybridisation probably also occurs but introgression simply does not persist - hinting at another instance of the environmental context dependence of introgression.

In the apple, radish and carrot cases highlighted above, the affected wild relatives still exist outside the areas where genetic assimilation occurred, demonstrating that genetic swamping is a localised phenomenon that may be restricted to specific areas and conditions. However, (repeated) genetic assimilation can lead to the local extinction of wild relatives, and therefore be a valid conservation concern, especially when it threatens fragmented, small or endemic populations.

6.2 Demographic swamping

Demographic swamping occurs when negative hybrid fitness effects lead to outbreeding depression and a subsequent drop in population size (Ellstrand, 2014). In the case of continuous gene flow, dwindling numbers may push the population into a negative spiral where high genetic drift makes it even harder to purge the deleterious variants through purifying selection. With each generation, as there are fewer and fewer individuals, the population is therefore less equipped to withstand the influx of deleterious alleles, and may ultimately enter the "extinction vortex" ending in population collapse and extinction (Bürger & Lynch, 1995;

Lynch et al., 1995). These risks are more pronounced in recipient populations that are small to begin with.

Originally described by Haygood et al. (2003), the concept of demographic swamping is closely tied to that of genetic swamping (see above), both of which result in reduced parent/pure wild population size and carry a risk of extinction, but through different mechanisms. A literature review of the connection between hybridisation and extinction found genetic swamping to be more frequent than demographic swamping, suggesting that extinction through assimilation is more likely than extinction through outbreeding depression *per se* (Todesco et al., 2016). A modelling study by Wolf et al. (2001) shows that this can happen remarkably fast (in less than five generations), especially in cases where there is no habitat differentiation between the hybrid and its parents. This suggests that adaption to different ecological conditions is key for preventing genetic assimilation of parent populations by their admixed offspring.

The two processes can also happen in conjunction, with demographic swamping accelerating the replacement of parent lineages, and genetic swamping leading to a loss of diversity and possible lower fitness over time. Only a couple of studies make claims of reduced genetic diversity and subsequent erosion of the wild gene pool (which are indicative of demographic swamping), both in connection to genetic swamping. This also shows that the two concepts are not always clearly distinguished in the literature. In pearl millet, genetic homogenisation of the wild gene pool has been linked to crop-to-wild admixture, raising concerns over the adaptive potential of the wild relative (Olodo et al., 2020). In wild yams, where nearly half of the individuals are recent or older hybrids, genetic swamping is considered to be ongoing and could lead to a loss of diversity (Scarcelli et al., 2017).

6.3 Genetic rescue

The opposite effect of demographic swamping can occur when the introgression of foreign alleles increases diversity and adaptive potential, leading to genetic rescue. While the discourse on crop-to-wild gene flow is dominated by concerns over reduced diversity, theoretical studies suggest that genetic rescue may be equally likely or even more likely than demographic swamping (Todesco et al., 2016). This process can be especially relevant for small populations that suffer from genetic load; for these populations, outcrossing with a separated population or species that is not too divergent can increase the effective population size across large parts of the genome and help to purge deleterious alleles.

Very few of the studies we surveyed explicitly measured the fitness of admixed individuals, although it has been suggested that the frequency of hybridisation and back-crossing in the European crab-apple *Malus sylvestris* (which is not the progenitor of domesticated apple) is a sign of positive fitness effects in these admixed individuals (Cornille et al., 2013), a hypothesis that was later confirmed in a regression analysis of fitness-related trait and cultivated ancestry proportion (Feurtey et al., 2017). A resurrection experiment of urban wild sunflowers showed that modern admixed individuals had a higher fitness than historical unadmixed individuals under modern climate conditions (Spear at al., 2023). While the exact regions responsible for these phenotypes were not identified, the combined evidence of crop-to-wild gene flow and fitness measurements in a common garden experiment suggests that the introgression of crop alleles may play a role in the genetic rescue of urban sunflowers exposed to climate stress.

Even though much of the original genomic diversity in crops is lost during the process of domestication, selection of specific traits may cause locally adapted crop varieties to harbour an enormous adaptive potential which could be of use for conservation of crop wild relatives, especially those that suffer from genetic erosion. Advantageous crop alleles that are already introgressed in wild populations are not just useful from a CWR conservation point of view; they could be useful for future breeding as well, if this diversity is ever lost from the cultivated gene pool (Ellstrand, 2018).

6.4 Adaptive introgression

Introgression is rarely ever only adaptive, but usually occurs locally against a background of global purging of deleterious alleles (Moran et al., 2021). Whereas genetic rescue implies an increase in fitness due to the overall reduction in genetic load (caused by an accumulation of weakly deleterious alleles), adaptive introgression occurs when there is positive selection on the introgression tract because the introgressed allele itself is adaptive. Demonstration of adaptive introgression therefore requires the confluence of several pieces of evidence: 1) introgression tracts need to be detected, 2) in a region of the genome that also carries a signature of selection, 3) which underlies relevant phenotypic variation, 4) that has a measurable fitness effect (Suarez-Gonzalez et al., 2018). Under the assumption of directional or positive selection, the combination of selection on an introgression tract leaves a special signature in the genome, as recombination breaks up an initially large introgressed haplotype into smaller blocks. This leaves a characteristic "volcano" shape where large regions of alleles at intermediate frequency surround a fixed adaptive allele (Setter et al., 2020). Given the background of global purging, adaptive introgression may be common even when levels of global admixture are low. This means that genome-wide measures of gene flow (such as the *D*-statistic or global ancestry) are not good measures for adaptive introgression, and that it can be overlooked if the right type and density of markers is not used. To detect adaptive introgression, the use of genome-wide SNP data, in combination with methods that can scan for introgression along the genome, is therefore advised.

Adaptive introgression is facilitated under specific conditions. Firstly, the globally deleterious nature of introgression requires that the fitness advantage of the allele(s) under selection is large enough that it can survive the initial phase of strong purging. The negative fitness effects of introgression may outweigh the positive fitness effects as hybridising species become more divergent, making it less likely for weakly beneficial alleles to persist (Edelman & Mallet, 2021). Hence, single loci of large effect are often favoured. Secondly, introgression can provide an important source of adaptive variation, especially in species that colonise new territory or are subject to environmental change (Edelman & Mallet, 2021). Adaptive introgression is therefore thought to be more common during range expansion or following the introduction of a plant into a novel environment. This has recently been observed in the wild progenitor of maize, which is showing signs of becoming invasive on the European mainland, in part due to introgression of key adaptive alleles from inbred maize cultivars (Le Corre et al., 2020), but may also be the case for other global weeds.

In extreme cases, when introgression is coupled with evolutionary innovations that allow the occupation of new niches, it can open up novel evolutionary pathways and pave the way for species formation and even adaptive radiations (Arnold & Kunte, 2017). Introgression of structural variants can also lead to the formation of new genetic architectures, for example when inversions suppress recombination and cause the evolution of cassettes of globally adaptive alleles that are inherited together, called "supergenes" (Jay et al., 2018). Most of the studies we surveyed focus on only two or a handful of species, precluding the identification of adaptive radiations. However, on a smaller and more recent time scale, we do see that introgression can lead to genetic differentiation of sub-populations and to the acquisition of novel traits that allow the wild or weedy relative to expand its range. This is most clearly visible in the case of weedy rice, which has a near global distribution and whose colonisation of new areas has been linked to crop-to-wild gene flow (Neik et al., 2019).

Some crop traits, especially those that have been selected precisely for their large effect loci, are ideal candidates for adaptive introgression into their wild relatives. We will go deeper into the acquisition of specific adaptive traits acquired by wild and weedy relatives below. An overview of traits currently being engineered in crop species with the help of NGTs is given in Appendix A (Table A2).

Resistance traits

Crops are often bred for improved biotic and abiotic resistance. In the case of apples, crop-to-wild gene flow has been implicated in conferring virus resistance (Feurtey et al., 2020) and aphid resistance (Denoirjean et al., 2021). In Europe, crop mutant alleles for photo-period sensitivity and herbicide resistance introgressed in the introduced maize wild relative *Zea mays* subsp. *mexicana* (Le Corre et al., 2020). Herbicide resistance has also been acquired by weedy rice in the United States (Wedger et al., 2022) and Argentina (Presotto et al., 2024). Examples in other species and other parts of the world probably exist, but were not included in our survey. Crop-to-wild introgression in urban sunflowers probably contributed to temperature and water responses and reduced plasticity in flowering time (Spear et al., 2023).

In some cases, introgression of crop alleles has led to the emergence of agricultural weeds or invasive species, which could be referred to as weedy or invasive (or both), depending on where they do most harm. Whereas weeds (or plant pests) are mostly considered an agricultural problem impacting production, invasive species can have profound impacts on natural ecosystems (Ellstrand et al., 2010). Although not all invasive species are weeds, and not all weeds are invasive, the two are closely related and both commonly descend from crop ancestors (Ellstrand et al., 2010). Schierenbeck & Ellstrand (2009) have described the mechanisms by which hybridisation can promote invasiveness, but crop-to-wild gene flow is not the only route to weedy or invasive characteristics. A large-scale study on the weedy relatives of Asian rice confirms that most weedy rice originated from local cultivars through a process of de-domestication or feralisation, without introgression (Qiu et al., 2020). While there are multiple processes leading to weediness - some involving crop-to-wild gene flow and others de-domestication or feralisation - the resulting agricultural weeds are all characterised by their ability to reproduce without human intervention. This makes them ideal vectors for the uncontrolled spread of crop genes into the natural environment. The presence of feral or weedy crop wild relatives can therefore be seen as a risk factor for crop-to-wild gene flow to more distant crop wild relatives.

Acquisition of the domestication syndrome

It is usually assumed that domestication alleles, especially those contributing to reproductive isolation, are less likely to be introgressed. But in at least one case, crop-to-wild introgression is claimed to have led to the transfer of domestication alleles (controlling shattering, hull colour, tillering and anthocyanin production) to a different wild (or pre-domesticated) population that was not the progenitor (Choi et al., 2017). This claim was later disputed on grounds of misinterpretation of the results (Civáň & Brown, 2018), showing just how difficult it can be to distinguish signatures of introgression from those left by selective sweeps. Choi & Purugganan (2018) subsequently reanalysed data originally published by Huang et al. (2012), disputing earlier claims that *japonica*, *indica* and *aus* were the result of three separate domestication events (Civáň et al., 2015). This exchange has become part of a lively scientific debate about the origins of rice that continues to spark new contributions, with ever increasing amounts of data and analysis techniques, including pan-genomes (Jing et al., 2023; Wu et al., 2023). While we did not encounter them in our survey, there are more crops with hypothesised multiple origins where introgression of crop alleles from one wild (or proto-domesticated) lineage to another may have catalysed evolution of the domestication syndrome, usually following the introduction of the earlier domesticate to a new area. One such example is the domestication of iron walnut (*Juglans sigillata*), which may have been facilitated by introgression from Persian walnut (*Juglans regia*) conferring, among others, an allele resulting in thinner nutshells (Ding et al., 2022).

6.5 Other evolutionary outcomes

Introgression of specific regions in the genome is just one possible outcome of a long-term process involving (recurrent) gene flow, selection, recombination and drift. The creation of a stable hybrid lineage, as described in the section on genetic assimilation, is another. The persistence of hybrids can also result in a hybrid swarm, consisting of individuals that are capable of backcrossing with their parent populations, or hybrid speciation, consisting of individuals that can intercross but are reproductively isolated from their parents. In populations with partially overlapping distributions, a hybrid zone can emerge where the two come into contact. In this case, gene flow may be recurrent, but restricted to a narrow geographic area - although some alleles could still spread throughout the parent populations if they are adaptive.

In addition to the direct consequences of crop-to-wild gene flow for the wild relatives, the impacts of crop-to-wild introgression could also extend beyond the crop wild relative itself as a result of ecological cascading effects, impacting other species in the ecosystem through biotic interactions on different trophic levels. Any positive or negative consequences on wild populations could therefore have unforeseen consequences that are difficult to predict.

7 Discussion and conclusion

7.1 Preliminary answers to research questions

From the available literature described above, we can provide some preliminary answers to the research questions outlined in the introduction.

How is introgression detected and monitored?

Firstly, we can see that there are tremendous opportunities for studying the presence, extent, direction, timing and mode of gene flow. The most advanced methods rely on genome-wide marker data such as SNPs and can lead to fine-grained insights on the regions that have introgressed, whether or not they are under selection, which lineages are affected by these signals and how long ago introgression took place. Despite these advances, few studies on crop-to-wild studies make full use of the methodological opportunities available, and many studies to this day still rely on microsatellite markers for population genetic inference. The most commonly used method to detect introgression is global ancestry, followed by *D*- and *f*-statistics and admixture graphs. Estimates of gene flow events or "pulses" were more common than estimates of continuous migration. Direction and timing were rarely explicitly tested, although extent (whether in the genome or in populations) was frequently quantified. The most powerful full-likelihood and Bayesian models, capable of estimating many demographic parameters simultaneously, were the least commonly applied.

How common is crop-to-wild introgression in different crop species?

Secondly, crop-to-wild gene flow appears to be incredibly common, as not a single study we surveyed reported no evidence for it at all. As explained before, this might partially reflect publication and reporting biases. However, we observed no clear trend in the levels of introgression reported across taxa, and we found both low and extremely high estimates of admixture levels, gene flow and migration in the studies we surveyed. Regardless of how much introgression was found, its almost universal presence among the study systems analysed is consistent with the growing consensus that introgression happens more frequently than long thought. Reviews and meta-analyses conservatively estimate that about 10% of (wild) plant species undergo homoploid hybridisation (Yakimowski & Rieseberg, 2014) and that interspecific introgression occurs in about 25% of plant species (Mallet, 2005). This only concerns gene flow between species, whereas, on an evolutionary time scale, many crops only diverged very recently from their ancestor and are therefore not considered properly different species. Hence, crop-to-wild gene flow with close relatives is probably even more common. In the surveyed articles, we see introgression from crops both into their progenitor species and other wild relatives, with no clear differences in the observed frequency or extent between more closely and distantly related species. In fact, admixture sometimes attained higher frequencies in the population of distant relatives than in the wild progenitor (Cornille et al., 2013) - although this may be explained to some extent by the difficulty of inferring gene flow between close relatives, who share ancestry across much of their genome.

What factors and mechanisms determine the possibility and extent of crop-to-wild introgression?

Thirdly, the factors that influence the prevalence of crop-to-wild introgression are a mix of pre- and post-zygotic barriers, that vary by crop, geographic location and agricultural or environmental context. In general, pre-zygotic barriers that prevent fertilisation contribute more towards reproductive isolation than post-zygotic barriers that impact hybrid viability, fertility and fitness, and they tend to be weaker in perennial species than in annual species. Sympatry, overlapping flowering times and, in the case of animal vectors, shared pollinators are all important conditions for gene flow to occur. Crop-to-wild gene flow is more likely to occur at shorter distances between cultivated and wild populations and at higher crop densities. Human-mediated seed dispersal, plant introductions and climate change all affect species distributions (and sometimes flowering times) and increasingly facilitate crop-to-wild gene flow outside their native range. While outcrossing species are more likely to hybridise, gene flow is even observed between predominantly selfing species. Similarly, genome characteristics once thought to guarantee reproductive isolation, such as differences in ploidy, don't always prevent gene flow. Genetic divergence does increase the likelihood of genetic incompatibilities and negative fitness effects, but to what extent these ultimately decide the fate of introgression tracts following hybridisation depends on a complicated mix of selection, recombination and demographic factors. Positive selection will favour the retention of introgression tracts, whereas negative selection will lead to purging, but the effectiveness of this process is affected by both linkage disequilibrium and genetic drift. Predicting the trajectory of introgression is therefore notoriously difficult, as it depends on the number and genetic composition of the parental individuals contributing the hybridisation event (also referred to as founder effects) and the environmental context in which it happens (Moran et al, 2021).

What consequences does crop-to-wild introgression have for natural populations of crop wild relatives?

Fourthly, while the effects of introgression are presumed to be globally deleterious (across the genome), they may be locally adaptive. Introgression tracts are preferentially retained when they are neutral or under positive selection. Following an initial phase of strong selection against large haplotypes with many linked alleles (some or many of which deleterious in nature), introgression tracts break down in size due to recombination and their fitness effects become smaller, allowing positive selection to favour beneficial alleles and successive generations of hybrids and backcrosses to move closer to their phenotypic optimum. This means that genomic extent is a poor indicator of the phenotypic consequences of introgression: the majority of deleterious alleles may be purged, whereas small regions harbouring adaptive alleles may be retained. Genomic regions with a low recombination rate, and populations with a small effective population size are less capable of purging deleterious alleles, and hence more susceptible to maladaptation. In these cases, outbreeding depression and ultimately dwindling population sizes (referred to as demographic swamping) is a risk.

In contrast to loss of genetic diversity or even extinction due to demographic swamping, when introgression is neutral or beneficial it can lead to genetic assimilation of the parent populations, where the original populations are gradually replaced by admixed individuals. This seems to be more common than demographic swamping, and has been reported in a number of studies surveyed, even in the absence of positive selection. Introgression can also "rescue" vulnerable populations by increasing their diversity and lifting inbreeding depression. The relative likelihood of genetic rescue versus outbreeding depression depends on the original size and genetic load of the donor and recipient populations, and the divergence time (which is a rough proxy for the number of genetic incompatibilities) between them. Tied to genetic rescue, but slightly different, is adaptive introgression, where the introgressed alleles themselves are under positive selection. This was the second most observed effect, in addition to genetic assimilation. The occurrence of adaptive introgression can be hardly detectable in genome-wide estimates, but have far-ranging phenotypic consequences, often tied to a single locus with large effect. The most commonly reported phenotypes conferred by adaptive introgression were herbicide resistance, followed by biotic resistance against pests and pathogens, and flowering time. These all impact the fitness of the wild relative, potentially leading to the emergence of novel weeds or invasive plants, and could further break down reproductive barriers.

How can the extent of crop-to-wild introgression and any adverse consequences it might have for wild populations be mitigated?

In line with the view that pre-zygotic barriers to reproduction tend to be stronger than post-zygotic barriers, gene flow is often contained by attempting to prevent fertilisation, for example by delaying flowering time of crops (Kwit et al., 2011). Other mitigation options that introduce pre-zygotic barriers to prevent the spread of genetically engineered alleles include male sterility or the excision of the engineered allele from pollen (Stewart et al., 2003). However, insofar as these measures are limited pollen-producing crops, they don't prevent fertilisation of cultivated plants by wild pollen, which has been reported in several of the studies we surveyed (Muller et al., 2011; Scarcelli et al., 2017). This direction of gene flow could to some extent be controlled by cleistogamy, which leads to flowers that don't open (Kwit et al., 2011). Cleistogamy thus enforces self-pollination and ensures that crop flowers can't be pollinated by wild relatives, but the seeds of cultivated species could still be recruited in the natural environment. The total prevention of flower formation would circumvent this issue, but is only a viable option for non-seed products (Chandler & Dunwell, 2008).

Mitigation measures leveraging post-zygotic barriers include the use of ploidy barriers or other genetic incompatibilities (Kwit et al., 2011). Suppression of recombination around the engineered allele is another option; in addition, it is possible to engineer alleles in areas of low recombination/high linkage disequilibrium (which are less likely to cross over into other chromosomes), or alternatively in non-nuclear genomes such as the chloroplast, which are maternally inherited and therefore not liable to escape via pollen-mediated gene flow (Stewart et al., 2003). The latter again is not resistant to seed-mediated gene flow. Post-zygotic mitigation strategies based on reduced hybrid fitness include coupling the engineered allele with a mitigation gene that is beneficial or neutral in the crop, but deleterious in the weedy or wild recipient; or the use of selectively terminable lines that are sensitive to specific herbicides (Kwit et al., 2011).

Many of these methods, both pre- and post-zygotic, themselves rely on further genetic engineering in addition to any traits of interest that might be engineered, and none of them are fool proof. Non-molecular methods relying on geographic isolation could be the most effective option when NGT plants are grown only in areas where their relatives almost certainly don't exist (Chandler & Dunwell, 2008). However, as demonstrated in the case of maize and sunflower, these relatives could still be introduced, causing unforeseen problems in the future (Le Corre et al., 2020; Roumet et al., 2013).

7.2 The risk equation: likelihood and impact

The answers to the research questions collectively show that crop-to-wild introgression is common in many types of crops; that it occurs even under conditions where it is not expected; that it has diverse consequences; and that it is almost impossible to prevent. With reference to NGT plants, where the possibility of gene flow over short time scales is a relevant factor for risk assessments, the studies surveyed here do not always give a good indication of current risk of (rapid) introgression. Some phylogenomic studies, for example on strawberry (Fan & Whitaker, 2024), barley (Jin et al., 2024) and *Brassica* (Saban et al., 2023), inferred historical gene flow, but did not necessarily have the geographic or temporal coverage to be able to ascertain whether gene flow happened recently or is still ongoing. Exceptions to this are studies analysing many individuals of wild relatives, often with fewer markers (e.g. in apple or lettuce), or those using time series data to track the increase in admixture in real time (e.g. in maize or sunflower).

Some methods and sampling designs are better suited to pick up signs of recent or ongoing introgression. These considerations are described in more detail below (Monitoring), but in general: detecting recent introgression requires fewer markers and denser population sampling than older introgression. In these cases, information from allele frequencies or site patterns (such as D_{FS}) may suffice without requiring model-based inference of gene trees or phylogenetic networks - although care should always be taken to account for the possible confounding effects of population structure or selection. Nonetheless, even when phylogenomic studies have a deep time perspective at the expense of contemporary geographic and temporal resolution, evidence of historical introgression still demonstrates that 1) gene flow is possible, and 2) it could be found to be recurrent and more extensive when studied at the appropriate scale, even if that was not the explicit goal of the study at hand. In addition, what these studies do better than some others, is place the observed introgression in context: by demonstrating between which lineages gene flow occurred (and which are not involved), and which introgressed regions persist in the long-term versus the short-term.

The overall strong probability of crop-to-wild gene flow, which can be high when summed over large spatial and temporal scales, even when individual probabilities are low, means that mitigation strategies should be concerned with its potential long-term consequences, alongside preventative measures (Kwit et al., 2011). As mentioned before, in the context of NGTs, the consequences of crop-to-wild gene flow depend largely on the engineered traits. Loss of shattering or other domestication traits will most likely not confer fitness benefits on wild species outside an agricultural context (although they might be beneficial to weeds), while drought or pest tolerance might be adaptive. Adaptive alleles could have a positive conservation value for wild populations, by increasing the resilience of wild populations, but also pose a cost for society when they leads to the emergence of novel weeds or increase the vigour of existing ones.

These long-term consequences are uncertain in themselves and may differ from the early effects of gene flow, as selection pressures can change (meaning that advantageous alleles may become deleterious and vice versa), and drift can push some alleles to fixation whereas others disappear (Ellstrand, 2003a). At the same time, introgression can happen very rapidly, in a handful of generations, and some studies have reported remarkable differences over only a few decades (Olodo et al., 2020; Rojas-Barrera et al., 2019; Spear et al., 2023). The difficulties associated with predicting whether, and how much, introgression will happen and what its consequences will be, imply that more research is not necessarily guaranteed to provide better answers, and that attempting to eliminate uncertainty might be a futile endeavour (de Jong & Rong, 2013). Rather than chasing precise but unattainable risk estimates, resources might therefore be better spent on a conservative estimation of risk and careful monitoring.

The risks of crop-to-wild introgression are not limited to GMO or NGT crops, and exist for many traditionally bred cultivars. Most of the studies we surveyed did not explicitly study transgene introgression, but introgression in general, and the same consequences could arise regardless of how a trait was improved - through traditional means, genetic modification, or new genomic techniques. The risks of crop-to-wild introgression are, however, inflated by other kinds of human interference. Intensive agriculture and other human activities involving heavy plant traffic increase the likelihood of crop escape. Human-mediated seed dispersal can cause the introduction of (engineered) crops in areas where they are not cultivated or authorised (Rostoks et al., 2019). Climate change and other human impacts on the environment also alter the distribution of wild plants, forcing them outside their natural range and introducing them in areas where they could become invasive. All these factors render the development and production of NGT plants a cross-border problem and highlight the need for monitoring if the genetic material of these plants is to be contained.

7.3 Recommendations for monitoring and detection

While the studies we surveyed for this report all detected introgression between crops and their wild relatives, few of them explicitly addressed monitoring. The examples that come closest to monitoring are those that used time series to track changes in allele frequencies or admixture proportions over time (Olodo et al., 2020; Rojas-Barrera et al., 2019; Spear et al., 2023); only one of which linked introgression to phenotypic evolution (Spear et al., 2023). Other insights from the literature, however, offer some guidance on what an effective monitoring strategy could look like. Assuming that the baseline situation implies an absence of introgression, monitoring should aim to detect *early* signs of *recent* introgression. These signs will likely involve low numbers of individuals, who possess long introgression tracts; hence, broad sampling is desirable without the need for full genomic coverage. Depending on the range and vulnerability of the wild species, sampling could either be geographically representative, or focus on locations closer to the area of cultivation. The main challenge is to separate introgressed alleles from allele frequency changes due to standing variation. Hence, it is important to have a comprehensive overview of the genetic diversity contained in the wild versus cultivated gene pools, and understand their relationships and population structure. Low density marker data could be sufficient to identify early generation hybrids or backcrosses; however, if the intent is to monitor introgression along the genome, genome-wide markers should be coupled with appropriate scans for local introgression to verify which regions have introgressed or not. Given the costs associated with whole genome sequencing, this might not always be feasible. When specific regions of the genome are of interest, targeted capture and sequencing could be a suitable alternative to enable more extensive monitoring than genome-wide scans of introgression, at the cost of being able to compare these regions to genome-wide signals.

The ideal conditions for detecting and monitoring introgression are outlined in Box 3. These partially emerged from the literature, and were partially co-constructed with a panel of experts in population genetics and sustainable agriculture, during a workshop led by the authors of this report. As a best-case scenario, these criteria are not necessarily all practically attainable or likely to be fulfilled simultaneously. Pragmatic considerations should therefore guide choices in the type(s) of evidence gathered and identify permissible deviations from the optimal course, depending on the case concerned. For example, unlike traditional breeding and genetic modification, new genomic techniques allow for the introduction of targeted mutations in the genome, which could confer desirable (domestication) traits onto previously uncultivated wild plants. The introduced variants could be very small and would exist in a genetic background that is identical to the wild population, and therefore be difficult to distinguish from standing variation or natural variation. In this case, ensuring that introgression can properly be detected would require the simultaneous inheritance of multiple crop-derived alleles that can, in combination, be recognised as the unique result of gene flow. However, this would rely on the introduction of (controlled) off-target effects or longer diagnostic alleles, which would run counter to the rationale behind exempting NGTs from GMO regulations, which is precisely that they can be equivalent to or indistinguishable from the natural variation selected by traditional breeding techniques. The same could be said for some of the mitigation strategies outlined above: when more genetic engineering is required to monitor and mitigate the escape of engineered crop alleles, than it prevents from being released into nature, the question rises whether the remedy is possibly worse than the disease.

Box 3. The best-case scenario for detection and monitoring

A workshop conducted on 25 June with experts in crop population genetics and sustainable agriculture, from different research institutes in Montpellier (France), led to the identification of a theoretical list of criteria that evidence-based claims of crop-to-wild introgression should adhere to. These results suggest that credible claims regarding the risk of introgression from crops to wild relatives, should possess the following attributes:

1. **Data intensive**: have full-scale, genome-wide coverage at sufficient read depth and sampling size to reliably detect variants and allele frequencies - sufficient to ascertain haplotype structure at geographically representative population-level sampling.
2. **Demographically informed**: be based on knowledge of the genetic background and evolutionary history of (potentially) inter-crossing populations (incl. population structure, effective population size), in order to account for possible spurious results.
3. **Reliable and robust**: use (a combination of) the most robust methods that are appropriate for the type of data (e.g., SNPs of a certain quality and quantity) and the model system (e.g. ploidy, demographic history), and that minimise bias.
4. **Functionally relevant**: take into account the alleles and traits possibly affected.
5. **Context-dependent**: consider different phenotypic and fitness effects relative to genomic background (epistasis) and environmental context (gene-environment interaction)
6. **Theoretically sound**: be in line with predictions based on the latest knowledge of the (interplay between) relevant evolutionary forces and population dynamics influencing the system, generated by modelling/simulations
7. **Experimentally validated**: make use of positive and negative controls (in the case of NGTs also for on and off target effects) for detecting the source of a variant (and its phenotypic/fitness effect)

Ultimately the classification of (some) crop alleles engineered with NGTs as equivalent to the outcomes of traditional breeding is a political decision, and exposes underlying perceptions regarding the acceptability of and risks associated with these technologies compared to conventional genetic modification. However, legalisation of NGT crops for the market by the EU does not mean that other stakeholders are absolved from thinking about the potential implications of their application. Society still has a responsibility to consider where and under which conditions NGT crops are desirable and acceptable, and how to deal with the consequences, should introgression occur. Data collection and analysis of (genetic) field data is time and resource intensive and these resources are by definition limited. This requires us to think about where and when resources to detect and monitor introgression should be allocated.

Where is introgression most likely to occur? Where are the possible consequences the most severe? Knowing where introgression is almost certain to happen, where the consequences would be unacceptable, or where the resources for detection and monitoring simply do not exist, will allow us to set the socio-political boundaries for the application of NGTs.

7.4 Future perspectives

This review has attempted to summarise the scientific progress that has been made in the study and detection of crop-to-wild introgression over the past decade and a half. The surveyed studies expose a rich and varied body of literature, encompassing a wide range of crop systems, sequencing techniques, analytical methods and disciplinary angles. Given our strict eligibility criteria for inclusion of literature, more evidence likely exists that we did not encounter, both on species that were included in this survey and those that were not. Follow-up studies could therefore conduct more thorough reviews for specific species of interest, adopting a lower specificity and higher sensitivity approach. The target genera identified in this report (Table A1) can serve as a guide for prioritising taxa that are relevant for a Dutch or European context and that warrant more research. This research may expose additional insights and methodologies not surveyed here.

While significant methodological advances have been made in the last decade and a half, both in the realm of sequencing technologies and in the realm of population genetic inference, not all of the current state-of-the-art techniques are routinely applied (yet) to the case of crops and their wild relatives. This strongly suggests that more fine-scale and reliable evidence on the prevalence and extent of crop-to-wild introgression can be gained, and is achievable pending sufficient funding and research interest. The use of genome-wide data and explicit tests of introgression, combining complementary statistics that each address a different aspect of gene flow (whether direction, timing or extent), is likely to yield a more complete picture of how often, when and how much introgression occurs in different crop systems. The application of more advanced statistics is, however, unchangeably sensitive to sampling choices, and hence caution remains to be exercised in selecting the lineages of interest, defining their relationships, testing their demographic history, and interpreting the results.

Despite methodological limitations, the available evidence strongly suggests that even if the prevalence of crop-to-wild introgression is difficult to quantify objectively, its likelihood is often non-zero. Moreover, the patterns observed are not static, but temporally dynamic and context dependent. The near certainty of introgression, coupled with the analytical contingencies that complicate predictions of its ultimate extent, means that quantifying and minimising the likelihood of introgression anywhere in the genome is probably less meaningful than trying to minimise the consequences if it does happen. These consequences ultimately depend on the traits affected, and will be most noticeable in the case of adaptive introgressions that confer a fitness advantage on wild or weedy relatives, with possible negative outcomes for nature and agriculture. The risk of using NGTs for crop engineering therefore lies not so much in the "exact" probability of introgression (which is dependent on so many factors that it is per definition shrouded in uncertainty), as in the specific alleles targeted and their phenotypic effects in wild populations. This suggests that future risk assessment on NGT crops should shift their attention from attaining precise estimates of the likelihood that introgression will occur, to mapping its possible impacts. A good starting point for this would be the species and traits currently being engineered in global research and development projects using NGTs (Table A2).

The interconnected fields of plant molecular breeding and genomics continue to evolve at a rapid pace, and future empirical studies on crop-to-wild introgression will probably benefit from ever-growing amounts of data, increasingly sophisticated analytical tools and expanding computational resources. As sequencing technologies become cheaper and more accurate, and statistical methods for data analysis more powerful, future advances in this domain will probably make more and more use of deep learning and artificial intelligence. While this offers exciting prospects for detecting complex signatures of selection and introgression more reliably and effectively, the study of crop-to-wild introgression will also benefit from continued theoretical developments that aim to disentangle the interplay of the various evolutionary processes that ultimately determine the pattern of introgression that is observed. Both sophisticated models for pattern recognition and an improved mechanistic understanding of the underlying processes will help to propel the study of introgression forward, and ultimately aid in the detection, monitoring and management of crop-to-wild gene flow.

References

Adugna, A., & Bekele, E. (2017). Indirect estimates reveal the potential of transgene flow in the crop-wild-weed Sorghum *bicolor* complex in its centre of origin, Ethiopia. *PLANT GENETIC RESOURCES-CHARACTERIZATION AND UTILIZATION*, *15*(6), 496–505. https://doi.org/10.1017/S1479262116000204

Ahmad, A., Jamil, A., & Munawar, N. (2023). GMOs or non-GMOs? The CRISPR Conundrum. *Frontiers in Plant Science*, *14*. https://doi.org/10.3389/fpls.2023.1232938

Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, *19*(9), 1655–1664. https://doi.org/10.1101/gr.094052.109

Andersson, M. S., & de Vicente, M. C. (2010). *Gene Flow between Crops and Their Wild Relatives*. The Johns Hopkins University Press.

Anses. (2024). *Risks and socio-economic issues related to NTG plants.* (2021-SA-0019). French National Health and Safety Agency of Food, Environment and Labour. https://www.anses.fr/fr/system/files/BIORISK2021SA0019EN.pdf

Armstrong, T. T., Fitzjohn, R. G., Newstrom, L. E., Wilton, A. D., & Lee, W. G. (2005). Transgene escape: What potential for crop-wild hybridization? *Molecular Ecology*, *14*(7), 2111–2132. Scopus. https://doi.org/10.1111/j.1365-294X.2005.02572.x

Arnold, M. L., & Kunte, K. (2017). Adaptive Genetic Exchange: A Tangled History of Admixture and Evolutionary Innovation. *Trends in Ecology & Evolution*, *32*(8), 601–611. https://doi.org/10.1016/j.tree.2017.05.007

Arrigo, N., Guadagnuolo, R., Lappe, S., Pasche, S., Parisod, C., & Felber, F. (2011). Gene flow between wheat and wild relatives: Empirical evidence from Aegilops geniculata, Ae. Neglecta and Ae. Triuncialis. *EVOLUTIONARY APPLICATIONS*, *4*(5), 685–695. https://doi.org/10.1111/j.1752-4571.2011.00191.x

Barchi, L., Aprea, G., Rabanus-Wallace, M. T., Toppino, L., Alonso, D., Portis, E., Lanteri, S., Gaccione, L., Omondi, E., van Zonneveld, M., Schafleitner, R., Ferrante, P., Börner, A., Stein, N., Díez, M. J., Lefebvre, V., Salinier, J., Boyaci, H. F., Finkers, R., … Giuliano, G. (2023). Analysis of >3400 worldwide eggplant accessions reveals two independent domestication events and multiple migration-diversification routes. *The Plant Journal*, *116*(6), 1667–1680. https://doi.org/10.1111/tpj.16455

Bartolić, P., Morgan, E. J., Padilla-García, N., & Kolář, F. (2024). Ploidy as a leaky reproductive barrier: Mechanisms, rates and evolutionary significance of interploidy gene flow. *Annals of Botany*, *134*(4), 537–550. https://doi.org/10.1093/aob/mcae096

Baute, G., Owens, G., Bock, D., & Rieseberg, L. (2016). Genome-wide genotyping-by-sequencing data provide a high-resolution view of wild Helianthus diversity, genetic structure, and interspecies gene flow. *AMERICAN JOURNAL OF BOTANY*, *103*(12), 2170–2177. https://doi.org/10.3732/ajb.1600295

Beckman, N. G., Bullock, J. M., & Salguero-Gómez, R. (2018). High dispersal ability is related to fast life-history strategies. *Journal of Ecology*, *106*(4), 1349–1362. https://doi.org/10.1111/1365-2745.12989

Beerli, P., & Felsenstein, J. (2001). Maximum likelihood estimation of a migration matrix and effective population sizes in n subpopulations by using a coalescent approach. *Proceedings of the National Academy of Sciences*, *98*(8), 4563–4568. https://doi.org/10.1073/pnas.081068098

Berg, P., Baltimore, D., Boyer, H. W., Cohen, S. N., Davis, R. W., Hogness, D. S., Nathans, D., Roblin, R., Watson, J. D., Weissman, S., & Zinder, N. D. (1974). Potential Biohazards of Recombinant DNA Molecules. *Science*, *185*(4148), 303–303. https://doi.org/10.1126/science.185.4148.303

Bezdek, J. C. (1981). *Pattern Recognition with Fuzzy Objective Function Algorithms*. Plenum Press.

Bina, H., Yousefzadeh, H., Venon, A., Remoué, C., Rousselet, A., Falque, M., Faramarzi, S., Chen, X., Samanchina, J., Gill, D., Kabaeva, A., Giraud, T., Hosseinpour, B., Abdollahi, H., Gabrielyan, I., Nersesyan, A., & Cornille, A. (2022). Evidence of an additional centre of apple domestication in Iran, with contributions from the Caucasian crab apple Malus orientalis Uglitzk. To the cultivated apple gene pool. *MOLECULAR ECOLOGY*, *31*(21), 5581–5601. https://doi.org/10.1111/mec.16667

Blischak, P. D., Chifman, J., Wolfe, A. D., & Kubatko, L. S. (2018). HyDe: A Python Package for Genome-Scale Hybridization Detection. *Systematic Biology*, *67*(5), 821–829. https://doi.org/10.1093/sysbio/syy023

Brandenburg, J.-T., Mary-Huard, T., Rigaill, G., Hearne, S. J., Corti, H., Joets, J., Vitte, C., Charcosset, A., Nicolas, S. D., & Tenaillon, M. I. (2017). Independent introductions and admixtures have contributed to adaptation of European maize and its American counterparts. *PLOS Genetics*, *13*(3), e1006666. https://doi.org/10.1371/journal.pgen.1006666

Bruetschy, C. (2019). The EU regulatory framework on genetically modified organisms (GMOs). *Transgenic Research*, *28*(2), 169–174. https://doi.org/10.1007/s11248-019-00149-y

Burgarella, C., Barnaud, A., Kane, N. A., Jankowski, F., Scarcelli, N., Billot, C., Vigouroux, Y., & Berthouly-Salazar, C. (2019). Adaptive Introgression: An Untapped Evolutionary Mechanism for Crop Adaptation. *Frontiers in Plant Science*, *10*. https://doi.org/10.3389/fpls.2019.00004

Bürger, R., & Lynch, M. (1995). EVOLUTION AND EXTINCTION IN A CHANGING ENVIRONMENT: A QUANTITATIVE-GENETIC ANALYSIS. *Evolution*, *49*(1), 151–163. https://doi.org/10.1111/j.1558-5646.1995.tb05967.x

Buschbom, J., Yanbaev, Y., & Degen, B. (2011). Efficient Long-Distance Gene Flow into an Isolated Relict Oak Stand. *Journal of Heredity*, *102*(4), 464–472. https://doi.org/10.1093/jhered/esr023

Cai, L., Xi, Z., Lemmon, E. M., Lemmon, A. R., Mast, A., Buddenhagen, C. E., Liu, L., & Davis, C. C. (2021). The Perfect Storm: Gene Tree Estimation Error, Incomplete Lineage Sorting, and Ancient Gene Flow Explain the Most Recalcitrant Ancient Angiosperm Clade, Malpighiales. *Systematic Biology*, *70*(3), 491–507. https://doi.org/10.1093/SYSBIO/SYAA083

Chamberlain, S. A., & Szöcs, E. (2013). taxize: Taxonomic search and retrieval in R. *F1000Research*, *2*, 191–191. https://doi.org/10.12688/f1000research.2-191.v2

Chandler, S., & Dunwell, J. M. (2008). Gene Flow, Risk Assessment and the Environmental Release of Transgenic Plants. *Critical Reviews in Plant Sciences*, *27*(1), 25–49. https://doi.org/10.1080/07352680802053916

Chao, C. T., & Krueger, R. R. (2007). The Date Palm (Phoenix dactylifera L.): Overview of Biology, Uses, and Cultivation. *HortScience*, *42*(5), 1077–1082. https://doi.org/10.21273/HORTSCI.42.5.1077

Chen, L., Wolf, A. B., Fu, W., Li, L., & Akey, J. M. (2020). Identifying and Interpreting Apparent Neanderthal Ancestry in African Individuals. *Cell*, *180*(4), 677-687.e16. https://doi.org/10.1016/j.cell.2020.01.012

Chen, X., Avia, K., Forler, A., Remoué, C., Venon, A., Rousselet, A., Lucas, G., Kwarteng, A. O., Rover, R., Le Guilloux, M., Belcram, H., Combes, V., Corti, H., Olverà-Vazquez, S., Falque, M., Alins, G., Kirisits, T., Ursu, T. M., Roman, A., … Cornille, A. (2023). Ecological and evolutionary drivers of phenotypic and genetic variation in the European crabapple [Malus sylvestris (L.) Mill.], a wild relative of the cultivated apple. *Annals of Botany*, *131*(6), 1025–1037. Scopus. https://doi.org/10.1093/aob/mcad061

Choi, J. Y., Platts, A. E., Fuller, D. Q., Hsing, Y.-I., Wing, R. A., Purugganan, M. D., & Kim, Y. (2017). The rice paradox: Multiple origins but single domestication in Asian Rice. *Molecular Biology and Evolution*, *34*(4), 969–979. Scopus. https://doi.org/10.1093/molbev/msx049

Choi, J. Y., & Purugganan, M. D. (2018). Multiple Origin but Single Domestication Led to Oryza sativa. *G3 Genes|Genomes|Genetics*, *8*(3), 797–803. https://doi.org/10.1534/g3.117.300334

Christie, K., Fraser, L. S., & Lowry, D. B. (2022). The strength of reproductive isolating barriers in seed plants: Insights from studies quantifying premating and postmating reproductive barriers over the past 15 years. *Evolution*, *76*(10), 2228–2243. https://doi.org/10.1111/evo.14565

Civáň, P., & Brown, T. A. (2018). Role of genetic introgression during the evolution of cultivated rice (Oryza sativa L.). *BMC Evolutionary Biology*, *18*(1), 57. https://doi.org/10.1186/s12862-018-1180-7

Civáň, P., Craig, H., Cox, C. J., & Brown, T. A. (2015). Three geographically separate domestications of Asian rice. *Nature Plants*, *1*(11), 1–5. https://doi.org/10.1038/nplants.2015.164

Cornille, A., Feurtey, A., Gélin, U., Ropars, J., Misvanderbrugge, K., Gladieux, P., & Giraud, T. (2015). Anthropogenic and natural drivers of gene flow in a temperate wild fruit tree: A basis for conservation and breeding programs in apples. *Evolutionary Applications*, *8*(4), 373–384. https://doi.org/10.1111/eva.12250

Cornille, A., Giraud, T., Smulders, M. J. M., Roldán-Ruiz, I., & Gladieux, P. (2014). The domestication and evolutionary ecology of apples. *Trends in Genetics*, *30*(2), 57–65. https://doi.org/10.1016/j.tig.2013.10.002

Cornille, A., Gladieux, P., & Giraud, T. (2013). Crop-to-wild gene flow and spatial genetic structure in the closest wild relatives of the cultivated apple. *EVOLUTIONARY APPLICATIONS*, *6*(5), 737–748. https://doi.org/10.1111/eva.12059

Cornuet, J.-M., Pudlo, P., Veyssier, J., Dehne-Garcia, A., Gautier, M., Leblois, R., Marin, J.-M., & Estoup, A. (2014). DIYABC v2.0: A software to make approximate Bayesian computation inferences about population history using single nucleotide polymorphism, DNA sequence and microsatellite data. *Bioinformatics*, *30*(8), 1187–1189. https://doi.org/10.1093/bioinformatics/btt763

Cornuet, J.-M., Santos, F., Beaumont, M. A., Robert, C. P., Marin, J.-M., Balding, D. J., Guillemaud, T., & Estoup, A. (2008). Inferring population history with DIY ABC: A user-friendly approach to approximate Bayesian computation. *Bioinformatics*, *24*(23), 2713–2719. https://doi.org/10.1093/bioinformatics/btn514

Dagilis, A. J., Peede, D., Coughlan, J. M., Jofre, G. I., D'Agostino, E. R. R., Mavengere, H., Tate, A. D., & Matute, D. R. (2021). *15 years of introgression studies: Quantifying gene flow across Eukaryotes* (p. 2021.06.15.448399). bioRxiv. https://doi.org/10.1101/2021.06.15.448399

de Jong, T., & Rong, J. (2013). Crop to wild gene flow: Does more sophisticated research provide better risk assessment? *ENVIRONMENTAL SCIENCE & POLICY*, *27*, 135–140. https://doi.org/10.1016/j.envsci.2012.12.002

Degnan, J. H. (2018). Modeling Hybridization Under the Network Multispecies Coalescent. *Systematic Biology*, *67*(5), 786–799. https://doi.org/10.1093/sysbio/syy040

Delplancke, M., Alvarez, N., Espíndola, A., Joly, H., Benoit, L., Brouck, E., & Arrigo, N. (2012). Gene flow among wild and domesticated almond species: Insights from chloroplast and nuclear markers. *EVOLUTIONARY APPLICATIONS*, *5*(4), 317–329. https://doi.org/10.1111/j.1752-4571.2011.00223.x

Denoirjean, T., Doury, G., Cornille, A., Chen, X., Hance, T., & Ameline, A. (2021). Genetic structure of Malus sylvestris and potential link with preference/performance by the rosy apple aphid pest Dysaphis plantaginea. *SCIENTIFIC REPORTS*, *11*(1). https://doi.org/10.1038/s41598-021-85014-x

Dias-Alves, T., Mairal, J., & Blum, M. G. B. (2018). Loter: A Software Package to Infer Local Ancestry for a Wide Range of Species. *Molecular Biology and Evolution*, *35*(9), 2318–2326. https://doi.org/10.1093/molbev/msy126

Dima, O., Custers, R., De Veirman, L., & Inzé, D. (2023). EU legal proposal for genome-edited crops hints at a science-based approach. *Trends in Plant Science*, *28*(12), 1350–1353. https://doi.org/10.1016/j.tplants.2023.09.014

Ding, Y.-M., Cao, Y., Zhang, W.-P., Chen, J., Liu, J., Li, P., Renner, S. S., Zhang, D.-Y., & Bai, W.-N. (2022). Population-genomic analyses reveal bottlenecks and asymmetric introgression from Persian into iron walnut during domestication. *Genome Biology*, *23*(1), 145. https://doi.org/10.1186/s13059-022-02720-z

Durand, E. Y., Patterson, N., Reich, D., & Slatkin, M. (2011). Testing for Ancient Admixture between Closely Related Populations. *Molecular Biology and Evolution*, *28*(8), 2239–2252. https://doi.org/10.1093/molbev/msr048

Eaton, D. A. R., & Ree, R. H. (2013). Inferring Phylogeny and Introgression using RADseq Data: An Example from Flowering Plants (Pedicularis: Orobanchaceae). *Systematic Biology*, *62*(5), 689–706. https://doi.org/10.1093/sysbio/syt032

Edelman, N. B., Frandsen, P. B., Miyagi, M., Clavijo, B., Davey, J., Dikow, R. B., García-Accinelli, G., Van Belleghem, S. M., Patterson, N., Neafsey, D. E., Challis, R., Kumar, S., Moreira, G. R. P., Salazar, C., Chouteau, M., Counterman, B. A., Papa, R., Blaxter, M., Reed, R. D., … Mallet, J. (2019). Genomic architecture and introgression shape a butterfly radiation. *Science*, *366*(6465), 594–599. https://doi.org/10.1126/science.aaw2090

Edelman, N. B., & Mallet, J. (2021). Prevalence and Adaptive Impact of Introgression. *Annual Review of Genetics*, *55*(Volume 55, 2021), 265–283. https://doi.org/10.1146/annurev-genet-021821-020805

Ellstrand, N. C. (2003a). Current knowledge of gene flow in plants: Implications for transgene flow. *PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B-BIOLOGICAL SCIENCES*, *358*(1434), 1163–1170. https://doi.org/10.1098/rstb.2003.1299

Ellstrand, N. C. (2003b). *Dangerous Liaisons? When Cultivated Plans Mate with Their Wild Relatives*. The Johns Hopkins University Press.

Ellstrand, N. C. (2014). Is gene flow the most important evolutionary force in plants? *American Journal of Botany*, *101*(5), 737–753. https://doi.org/10.3732/ajb.1400024

Ellstrand, N. C. (2018). Does introgression of crop alleles into wild and weedy living populations create cryptic in situ germplasm banks? *Molecular Ecology*, *27*(1), 38–40. Scopus. https://doi.org/10.1111/mec.14390

Ellstrand, N. C., Heredia, S. M., Leak-Garcia, J. A., Heraty, J. M., Burger, J. C., Yao, L., Nohzadeh-Malakshah, S., & Ridley, C. E. (2010). Crops gone wild: Evolution of weeds and invasives from domesticated ancestors. *Evolutionary Applications*, *3*(5–6), 494–504. https://doi.org/10.1111/j.1752-4571.2010.00140.x

Ellstrand, N. C., Meirmans, P., Rong, J., Bartsch, D., Ghosh, A., de Jong, T., Haccou, P., Lu, B., Snow, A., Stewart, C., Strasburg, J., van Tienderen, P., Vrieling, K., & Hooftman, D. (2013). Introgression of Crop Alleles into Wild or Weedy Populations. *ANNUAL REVIEW OF ECOLOGY, EVOLUTION, AND SYSTEMATICS, VOL 44*, *44*, 325–345. https://doi.org/10.1146/annurev-ecolsys-110512-135840

Ellstrand, N. C., & Rieseberg, L. H. (2016). When gene flow really matters: Gene flow in applied evolutionary biology. *Evolutionary Applications*, *9*(7), 833–836. https://doi.org/10.1111/eva.12402

Elworth, R. A. L., Ogilvie, H. A., Zhu, J., & Nakhleh, L. (2019). Advances in Computational Methods for Phylogenetic Networks in the Presence of Hybridization. In T. Warnow (Ed.), *Bioinformatics and Phylogenetics: Seminal Contributions of Bernard Moret* (pp. 317–360). Springer International Publishing. https://doi.org/10.1007/978-3-030-10837-3_13

Fan, Z., & Whitaker, V. M. (2024). Genomic signatures of strawberry domestication and diversification. *The Plant Cell*, *36*(5), 1622–1636. https://doi.org/10.1093/plcell/koad314

Feng, C., Wang, J., Liston, A., & Kang, M. (2023). Recombination Variation Shapes Phylogeny and Introgression in Wild Diploid Strawberries. *Molecular Biology and Evolution*, *40*(3), msad049. https://doi.org/10.1093/molbev/msad049

Feurtey, A., Cornille, A., Shykoff, J., Snirc, A., & Giraud, T. (2017). Crop-to-wild gene flow and its fitness consequences for a wild fruit tree: Towards a comprehensive conservation strategy of the wild apple in Europe. *EVOLUTIONARY APPLICATIONS*, *10*(2), 180–188. https://doi.org/10.1111/eva.12441

Feurtey, A., Guitton, E., De Gracia Coquerel, M., Duvaux, L., Shiller, J., Bellanger, M.-N., Expert, P., Sannier, M., Caffier, V., Giraud, T., Le Cam, B., & Lemaire, C. (2020). Threat to Asian wild apple trees posed by gene flow from domesticated apple trees and their “pestified” pathogens. *Molecular Ecology*, *29*(24), 4925–4941. Scopus. https://doi.org/10.1111/mec.15677

Flagel, L., Brandvain, Y., & Schrider, D. R. (2019). The Unreasonable Effectiveness of Convolutional Neural Networks in Population Genetic Inference. *Molecular Biology and Evolution*, *36*(2), 220–238. https://doi.org/10.1093/molbev/msy224

Forsythe, E. S., Nelson, A. D. L., & Beilstein, M. A. (2020). Biased Gene Retention in the Face of Introgression Obscures Species Relationships. *Genome Biology and Evolution*, *12*(9), 1646–1663. https://doi.org/10.1093/gbe/evaa149

Frichot, E., Mathieu, F., Trouillon, T., Bouchard, G., & François, O. (2014). Fast and Efficient Estimation of Individual Ancestry Coefficients. *Genetics*, *196*(4), 973–983. https://doi.org/10.1534/genetics.113.160572

Gao, H., Williamson, S., & Bustamante, C. D. (2007). A Markov Chain Monte Carlo Approach for Joint Inference of Population Structure and Inbreeding Rates From Multilocus Genotype Data. *Genetics*, *176*(3), 1635–1651. https://doi.org/10.1534/genetics.107.072371

Gaut, B. S., Díez, C. M., & Morrell, P. L. (2015). Genomics and the Contrasting Dynamics of Annual and Perennial Domestication. *Trends in Genetics*, *31*(12), 709–719. https://doi.org/10.1016/j.tig.2015.10.002

Geissler, C., Davidson, A., & Niesenbaum, R. A. (2023). The influence of climate warming on flowering phenology in relation to historical annual and seasonal temperatures and plant functional traits. *PeerJ*, *11*, e15188. https://doi.org/10.7717/peerj.15188

Gerber, S., Chadœuf, J., Gugerli, F., Lascoux, M., Buiteveld, J., Cottrell, J., Dounavi, A., Fineschi, S., Forrest, L. L., Fogelqvist, J., Goicoechea, P. G., Jensen, J. S., Salvini, D., Vendramin, G. G., & Kremer, A. (2014). High Rates of Gene Flow by Pollen and Seed in Oak Populations across Europe. *PLOS ONE*, *9*(1), e85130. https://doi.org/10.1371/journal.pone.0085130

Gower, G., Picazo, P. I., Fumagalli, M., & Racimo, F. (2021). Detecting adaptive introgression in human evolution using convolutional neural networks. *eLife*, *10*, e64669. https://doi.org/10.7554/eLife.64669

Green, R. E., Krause, J., Briggs, A. W., Maricic, T., Stenzel, U., Kircher, M., Patterson, N., Li, H., Zhai, W., Fritz, M. H.-Y., Hansen, N. F., Durand, E. Y., Malaspinas, A.-S., Jensen, J. D., Marques-Bonet, T., Alkan, C., Prüfer, K., Meyer, M., Burbano, H. A., … Pääbo, S. (2010). A Draft Sequence of the Neandertal Genome. *Science*, *328*(5979), 710–722. https://doi.org/10.1126/science.1188021

Gronau, I., Hubisz, M. J., Gulko, B., Danko, C. G., & Siepel, A. (2011). Bayesian inference of ancient human demography from individual genome sequences. *Nature Genetics*, *43*(10), 1031–1034. https://doi.org/10.1038/ng.937

Groot, M. H. M., Van de Wiel, C. C. M., Van Tienderen, P. H., & Den Nijs, H. C. M. (2003). *Hybridisation and introgression between crops and wild relatives: Current knowledge and research priorities in lieu of impending introductions of GM crops.* The Netherlands Commission on Genetic Modification (COGEM).

Guan, Y. (2014). Detecting Structure of Haplotypes and Local Ancestry. *Genetics*, *196*(3), 625–642. https://doi.org/10.1534/genetics.113.160697

Guo, Y., Himmelbach, A., Weiss, E., Stein, N., & Mascher, M. (2022). Six-rowed wild-growing barleys are hybrids of diverse origins. *PLANT JOURNAL*, *111*(3), 849–858. https://doi.org/10.1111/tpj.15861

Gutierrez, A., Carrera, A., Basualdo, J., Rodriguez, R., Cantamutto, M., & Poverene, M. (2010). Gene flow between cultivated sunflower and Helianthus petiolaris (Asteraceae). *EUPHYTICA*, *172*(1), 67–76. https://doi.org/10.1007/s10681-009-0045-y

Ha, Y., Oh, S., & Lee, S. (2021). Genetic Admixture in the Population of Wild Apple (Malus sieversii) from the Tien Shan Mountains, Kazakhstan. *GENES*, *12*(1). https://doi.org/10.3390/genes12010104

Hahn, M. W., & Hibbins, M. S. (2019). A Three-Sample Test for Introgression. *Molecular Biology and Evolution*, *36*(12), 2878–2882. https://doi.org/10.1093/molbev/msz178

Hamlin, J. A. P., Hibbins, M. S., & Moyle, L. C. (2020). Assessing biological factors affecting postspeciation introgression. *Evolution Letters*, *4*(2), 137–154. https://doi.org/10.1002/evl3.159

Haygood, R., Ives, A. R., & Andow, D. A. (2003). Consequences of recurrent gene flow from crops to wild relatives. *Proceedings of the Royal Society B: Biological Sciences*, *270*(1527), 1879–1886. Scopus. https://doi.org/10.1098/rspb.2003.2426

Hegde, S. G., Nason, J. D., Clegg, J. M., & Ellstrand, N. C. (2006). The Evolution of California's Wild Radish Has Resulted in the Extinction of Its Progenitors. *Evolution*, *60*(6), 1187–1197. https://doi.org/10.1111/j.0014-3820.2006.tb01197.x

Helmstetter, A., Oztolan-Erol, N., Lucas, S., & Buggs, R. (2020). Genetic diversity and domestication of hazelnut (Corylus avellana L.) in Turkey. *PLANTS PEOPLE PLANET*, *2*(4), 326–339. https://doi.org/10.1002/ppp3.10078

Heredia-Pech, M., Chávez-Pesqueira, M., Ortiz-García, M., Andueza-Noh, R., Chacón-Sánchez, M., & Martínez-Castillo, J. (2022). Consequences of introgression and gene flow on the genetic structure and diversity of Lima bean (Phaseolus lunatus L.) in its Mesoamerican diversity area. *PEERJ*, *10*. https://doi.org/10.7717/peerj.13690

Hernández, F., Palmieri, L., & Brunet, J. (2023). Introgression and persistence of cultivar alleles in wild carrot (Daucus carota) populations in the United States. *AMERICAN JOURNAL OF BOTANY*, *110*(11). https://doi.org/10.1002/ajb2.16242

Hibbins, M. S., & Hahn, M. W. (2019). The Timing and Direction of Introgression Under the Multispecies Network Coalescent. *Genetics*, *211*(3), 1059–1073. https://doi.org/10.1534/genetics.118.301831

Hibbins, M. S., & Hahn, M. W. (2022). Phylogenomic approaches to detecting and characterizing introgression. *Genetics*, *220*(2). https://doi.org/10.1093/GENETICS/IYAB173

Huang, X., Kurata, N., Wei, X., Wang, Z.-X., Wang, A., Zhao, Q., Zhao, Y., Liu, K., Lu, H., Li, W., Guo, Y., Lu, Y., Zhou, C., Fan, D., Weng, Q., Zhu, C., Huang, T., Zhang, L., Wang, Y., … Han, B. (2012). A map of rice genome variation reveals the origin of cultivated rice. *Nature*, *490*(7421), 497–501. https://doi.org/10.1038/nature11532

Hufford, M., Lubinksy, P., Pyhäjärvi, T., Devengenzo, M., Ellstrand, N., & Ross-Ibarra, J. (2013). The Genomic Signature of Crop-Wild Introgression in Maize. *PLOS GENETICS*, *9*(5). https://doi.org/10.1371/journal.pgen.1003477

Huynh, S., Marcussen, T., Felber, F., & Parisod, C. (2019). Hybridization preceded radiation in diploid wheats. *Molecular Phylogenetics and Evolution*, *139*. Scopus. https://doi.org/10.1016/j.ympev.2019.106554

Janzen, G. M., Wang, L., & Hufford, M. B. (2019). The extent of adaptive wild introgression in crops. *New Phytologist*, *221*(3), 1279–1288. https://doi.org/10.1111/nph.15457

Jay, P., Whibley, A., Frézal, L., Cara, M. Á. R. de, Nowell, R. W., Mallet, J., Dasmahapatra, K. K., & Joron, M. (2018). Supergene Evolution Triggered by the Introgression of a Chromosomal Inversion. *Current Biology*, *28*(11), 1839-1845.e3. https://doi.org/10.1016/j.cub.2018.04.072

Jenczewski, E., Ronfort, J., & Chèvre, A. M. (2003). Crop-to-wild gene flow, introgression and possible fitness effects of transgenes. *Environmental Biosafety Research*, *2*(1), 9–24. Scopus. https://doi.org/10.1051/ebr:2003001

Jiao, X., Flouri, T., & Yang, Z. (2021). Multispecies coalescent and its applications to infer species phylogenies and cross-species gene flow. *National Science Review*, *8*(12), 2021–2021. https://doi.org/10.1093/NSR/NWAB127

Jin, Y., Du, X., Jiang, C., Ji, W., & Yang, P. (2024). Disentangling sources of gene tree discordance for Hordeum species via target-enriched sequencing assays. *MOLECULAR PHYLOGENETICS AND EVOLUTION*, *199*. https://doi.org/10.1016/j.ympev.2024.108160

Jing, C.-Y., Zhang, F.-M., Wang, X.-H., Wang, M.-X., Zhou, L., Cai, Z., Han, J.-D., Geng, M.-F., Yu, W.-H., Jiao, Z.-H., Huang, L., Liu, R., Zheng, X.-M., Meng, Q.-L., Ren, N.-N., Zhang, H.-X., Du, Y.-S., Wang, X., Qiang, C.-G., … Ge, S. (2023). Multiple domestications of Asian rice. *Nature Plants*, *9*(8), 1221–1235. https://doi.org/10.1038/s41477-023-01476-z

Jombart, T., Devillard, S., & Balloux, F. (2010). Discriminant analysis of principal components: A new method for the analysis of genetically structured populations. *BMC Genetics*, *11*(1), 94. https://doi.org/10.1186/1471-2156-11-94

Karremans, A. P. (2024). A Historical Review of the Artificial Pollination of Vanilla planifolia: The Importance of Collaborative Research in a Changing World. *Plants*, *13*(22), Article 22. https://doi.org/10.3390/plants13223203

Korneliussen, T. S., Albrechtsen, A., & Nielsen, R. (2014). ANGSD: Analysis of Next Generation Sequencing Data. *BMC Bioinformatics*, *15*(1), 1–13. https://doi.org/10.1186/S12859-014-0356-4/TABLES/4

Kress, W. J., Soltis, D. E., Kersey, P. J., Wegrzyn, J. L., Leebens-Mack, J. H., Gostel, M. R., Liu, X., & Soltis, P. S. (2022). Green plant genomes: What we know in an era of rapidly expanding opportunities. *Proceedings of the National Academy of Sciences*, *119*(4), e2115640118. https://doi.org/10.1073/pnas.2115640118

Kwit, C., Moon, H. S., Warwick, S. I., & Stewart, C. N. (2011). Transgene introgression in crop relatives: Molecular evidence and mitigation strategies. *Trends in Biotechnology*, *29*(6), 284–293. https://doi.org/10.1016/j.tibtech.2011.02.003

Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczky, J., LeVine, R., McEwan, P., … The Wellcome Trust: (2001). Initial sequencing and analysis of the human genome. *Nature*, *409*(6822), 860–921. https://doi.org/10.1038/35057062

Lawson, D. J., van Dorp, L., & Falush, D. (2018). A tutorial on how not to over-interpret STRUCTURE and ADMIXTURE bar plots. *Nature Communications*, *9*(1), 3258. https://doi.org/10.1038/s41467-018-05257-7

Le Corre, V., Siol, M., Vigouroux, Y., Tenaillon, M., & Délye, C. (2020). Adaptive introgression from maize has facilitated the establishment of teosinte as a noxious weed in Europe. *PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA*, *117*(41), 25618–25627. https://doi.org/10.1073/pnas.2006633117

Lin, X., Fang, C., Liu, B., & Kong, F. (2021). Natural variation and artificial selection of photoperiodic flowering genes and their applications in crop adaptation. *aBIOTECH*, *2*(2), 156–169. https://doi.org/10.1007/s42994-021-00039-0

Lowry, D. B., Modliszewski, J. L., Wright, K. M., Wu, C. A., & Willis, J. H. (2008). The strength and genetic basis of reproductive isolating barriers in flowering plants. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*(1506), 3009–3021. https://doi.org/10.1098/rstb.2008.0064

Luikart, G., Kardos, M., Hand, B. K., Rajora, O. P., Aitken, S. N., & Hohenlohe, P. A. (2019). Population Genomics: Advancing Understanding of Nature. In O. P. Rajora (Ed.), *Population Genomics: Concepts, Approaches and Applications* (pp. 3–79). Springer International Publishing. https://doi.org/10.1007/13836_2018_60

Lynch, M., Conery, J., & Bürger, R. (1995). MUTATIONAL MELTDOWNS IN SEXUAL POPULATIONS. *Evolution*, *49*(6), 1067–1080. https://doi.org/10.1111/j.1558-5646.1995.tb04434.x

Malinsky, M., Matschiner, M., & Svardal, H. (2021). Dsuite—Fast D-statistics and related admixture evidence from VCF files. *Molecular Ecology Resources*, *21*(2), 584–595. https://doi.org/10.1111/1755-0998.13265

Malinsky, M., Svardal, H., Tyers, A. M., Miska, E. A., Genner, M. J., Turner, G. F., & Durbin, R. (2018). Whole-genome sequences of Malawi cichlids reveal multiple radiations interconnected by gene flow. *Nature Ecology & Evolution*, *2*(12), 1940–1955. https://doi.org/10.1038/s41559-018-0717-x

Mallet, J. (2005). Hybridization as an invasion of the genome. *Trends in Ecology & Evolution*, *20*(5), 229–237. https://doi.org/10.1016/j.tree.2005.02.010

Martin, S. H., & Amos, W. (2021). Signatures of Introgression across the Allele Frequency Spectrum. *Molecular Biology and Evolution*, *38*(2), 716–726. https://doi.org/10.1093/molbev/msaa239

Martin, S. H., Davey, J. W., & Jiggins, C. D. (2015). Evaluating the Use of ABBA–BABA Statistics to Locate Introgressed Loci. *Molecular Biology and Evolution*, *32*(1), 244–257. https://doi.org/10.1093/MOLBEV/MSU269

Meyer, R. S., & Purugganan, M. D. (2013). Evolution of crop species: Genetics of domestication and diversification. *Nature Reviews Genetics*, *14*(12), 840–852. https://doi.org/10.1038/nrg3605

Mondon, A., Owens, G., Poverene, M., Cantamutto, M., & Rieseberg, L. (2018). Gene flow in Argentinian sunflowers as revealed by genotyping-by-sequencing data. *EVOLUTIONARY APPLICATIONS*, *11*(2), 193–204. https://doi.org/10.1111/eva.12527

Monnet, F., Postel, Z., Touzet, P., Fraïsse, C., Peer, Y. V. de, Vekemans, X., & Roux, C. (2023). *Rapid establishment of species barriers in plants compared to animals* (p. 2023.10.16.562535). bioRxiv. https://doi.org/10.1101/2023.10.16.562535

Moran, B. M., Payne, C., Langdon, Q., Powell, D. L., Brandvain, Y., & Schumer, M. (2021). The genomic consequences of hybridization. *eLife*, *10*, e69016. https://doi.org/10.7554/eLife.69016

Muller, M., Latreille, M., & Tollon, C. (2011). The origin and evolution of a recent agricultural weed: Population genetic diversity of weedy populations of sunflower (Helianthus annuus L.) in Spain and France. *EVOLUTIONARY APPLICATIONS*, *4*(3), 499–514. https://doi.org/10.1111/j.1752-4571.2010.00163.x

Murase, K., Moriwaki, Y., Mori, T., Liu, X., Masaka, C., Takada, Y., Maesaki, R., Mishima, M., Fujii, S., Hirano, Y., Kawabe, Z., Nagata, K., Terada, T., Suzuki, G., Watanabe, M., Shimizu, K., Hakoshima, T., & Takayama, S. (2020). Mechanism of self/nonself-discrimination in Brassica self-incompatibility. *Nature Communications*, *11*(1), 4916. https://doi.org/10.1038/s41467-020-18698-w

Mutegi, E., Sagnard, F., Labuschagne, M., Herselman, L., Semagn, K., Deu, M., de Villiers, S., Kanyenji, B., Mwongera, C., Traore, P., & Kiambi, D. (2012). Local scale patterns of gene flow and genetic diversity in a crop-wild-weedy complex of sorghum (Sorghum bicolor (L.) Moench) under traditional agricultural field conditions in Kenya. *CONSERVATION GENETICS*, *13*(4), 1059–1071. https://doi.org/10.1007/s10592-012-0353-y

Neik, T., Chai, J., Tan, S., Sudo, M., Cui, Y., Jayaraj, J., Teo, S., Olsen, K., & Song, B. (2019). When West Meets East: The Origins and Spread of Weedy Rice Between Continental and Island Southeast Asia. *G3-GENES GENOMES GENETICS*, *9*(9), 2941–2950. https://doi.org/10.1534/g3.119.400021

O'Connor, K., Powell, M., Nock, C., & Shapcott, A. (2015). Crop to wild gene flow and genetic diversity in a vulnerable Macadamia (Proteaceae) species in New South Wales, Australia. *BIOLOGICAL CONSERVATION*, *191*, 504–511. https://doi.org/10.1016/j.biocon.2015.08.001

Olodo, K., Barnaud, A., Kane, N., Mariac, C., Faye, A., Couderc, M., Zekraouï, L., Dequincey, A., Diouf, D., Vigouroux, Y., & Berthouly-Salazar, C. (2020). Abandonment of pearl millet cropping and homogenization of its diversity over a 40 year period in Senegal. *PLOS ONE*, *15*(9). https://doi.org/10.1371/journal.pone.0239123

Olofsson, J. K., Tyler, T., Dunning, L. T., Hjertson, M., Rühling, Å., & Hansen, A. J. (2024). Morphological and genetic evidence suggest gene flow among native and naturalized mint species. *American Journal of Botany*, *111*(2), e16280. https://doi.org/10.1002/ajb2.16280

Omasheva, M., Flachowsky, H., Ryabushkina, N., Pozharskiy, A., Galiakparov, N., & Hanke, M. (2017). To what extent do wild apples in Kazakhstan retain their genetic integrity? *TREE GENETICS & GENOMES*, *13*(3). https://doi.org/10.1007/s11295-017-1134-z

Ouzzani, M., Hammady, H., Fedorowicz, Z., & Elmagarmid, A. (2016). Rayyan—A web and mobile app for systematic reviews. *Systematic Reviews*, *5*, 210. https://doi.org/10.1186/s13643-016-0384-4

Page, A., Gibson, J., Meyer, R. S., & Chapman, M. A. (2019). Eggplant Domestication: Pervasive Gene Flow, Feralization, and Transcriptomic Divergence. *Molecular Biology and Evolution*, *36*(7), 1359–1372. https://doi.org/10.1093/molbev/msz062

Pajkovic, M., Lappe, S., Barman, R., Parisod, C., Neuenschwander, S., Goudet, J., Alvarez, N., Guadagnuolo, R., Felber, F., & Arrigo, N. (2014). Wheat alleles introgress into selfing wild relatives: Empirical estimates from approximate Bayesian computation in Aegilops triuncialis. *MOLECULAR ECOLOGY*, *23*(20), 5089–5101. https://doi.org/10.1111/mec.12918

Parisi, C., & Rodríguez-Cerezo, E. (2021). *Current and future market applications of new genomic techniques*. Publications Office of the European Union. https://data.europa.eu/doi/10.2760/02472

Parisod, C., Definod, C., Sarr, A., Arrigo, N., & Felber, F. (2013). Genome-specific introgression between wheat and its wild relative Aegilops triuncialis. *JOURNAL OF EVOLUTIONARY BIOLOGY*, *26*(1), 223–228. https://doi.org/10.1111/jeb.12040

Patterson, N., Moorjani, P., Luo, Y., Mallick, S., Rohland, N., Zhan, Y., Genschoreck, T., Webster, T., & Reich, D. (2012). Ancient Admixture in Human History. *Genetics*, *192*(3), 1065–1093. https://doi.org/10.1534/genetics.112.145037

Pease, J. B., & Hahn, M. W. (2015). Detection and Polarization of Introgression in a Five-Taxon Phylogeny. *Systematic Biology*, *64*(4), 651–662. https://doi.org/10.1093/sysbio/syv023

Peter, B. M. (2016). Admixture, Population Structure, and F-Statistics. *Genetics*, *202*(4), 1485–1501. https://doi.org/10.1534/genetics.115.183913

Pfeifer, B., & Kapan, D. D. (2019). Estimates of introgression as a function of pairwise distances. *BMC Bioinformatics*, *20*(1), 207. https://doi.org/10.1186/s12859-019-2747-z

Pickrell, J. K., & Pritchard, J. K. (2012). Inference of Population Splits and Mixtures from Genome-Wide Allele Frequency Data. *PLOS Genetics*, *8*(11), e1002967. https://doi.org/10.1371/journal.pgen.1002967

Presotto, A., Hernández, F., Vercellino, R. B., Kruger, R. D., Fontana, M. L., Ureta, M. S., Crepy, M., Auge, G., & Caicedo, A. (2024). Introgression from local cultivars is a driver of agricultural adaptation in Argentinian weedy rice. *Molecular Ecology*, *33*(11). Scopus. https://doi.org/10.1111/mec.17368

Price, A. L., Tandon, A., Patterson, N., Barnes, K. C., Rafaels, N., Ruczinski, I., Beaty, T. H., Mathias, R., Reich, D., & Myers, S. (2009). Sensitive Detection of Chromosomal Segments of Distinct Ancestry in Admixed Populations. *PLOS Genetics*, *5*(6), e1000519. https://doi.org/10.1371/journal.pgen.1000519

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of Population Structure Using Multilocus Genotype Data. *Genetics*, *155*(2), 945–959. https://doi.org/10.1093/genetics/155.2.945

Qiu, J., Jia, L., Wu, D., Weng, X., Chen, L., Sun, J., Chen, M., Mao, L., Jiang, B., Ye, C., Turra, G. M., Guo, L., Ye, G., Zhu, Q.-H., Imaizumi, T., Song, B.-K., Scarabel, L., Merotto, A., Jr, Olsen, K. M., & Fan, L. (2020). Diverse genetic mechanisms underlie worldwide convergent rice feralization. *Genome Biology*, *21*(1). Scopus. https://doi.org/10.1186/s13059-020-01980-x

Rabanus-Wallace, M. T., Hackauf, B., Mascher, M., Lux, T., Wicker, T., Gundlach, H., Baez, M., Houben, A., Mayer, K. F. X., Guo, L., Poland, J., Pozniak, C. J., Walkowiak, S., Melonek, J., Praz, C. R., Schreiber, M., Budak, H., Heuberger, M., Steuernagel, B., … Stein, N. (2021). Chromosome-scale genome assembly provides insights into rye biology, evolution and agronomic potential. *Nature Genetics*, *53*(4), 564–573. Scopus. https://doi.org/10.1038/s41588-021-00807-0

Raj, A., Stephens, M., & Pritchard, J. K. (2014). fastSTRUCTURE: Variational Inference of Population Structure in Large SNP Data Sets. *Genetics*, *197*(2), 573–589. https://doi.org/10.1534/genetics.114.164350

Rojas-Barrera, I., Wegier, A., González, J., Owens, G., Rieseberg, L., & Piñero, D. (2019). Contemporary evolution of maize landraces and their wild relatives influenced by gene flow with modern maize varieties. *PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA*, *116*(42), 21302–21311. https://doi.org/10.1073/pnas.1817664116

Rostoks, N., Grantiņa-Ieviņa, L., Ieviņa, B., Evelone, V., Valciņa, O., & Aleksejeva, I. (2019). Genetically modified seeds and plant propagating material in Europe: Potential routes of entrance and current status. *Heliyon*, *5*(2), e01242. https://doi.org/10.1016/j.heliyon.2019.e01242

Roumet, M., Noilhan, C., Latreille, M., David, J., & Muller, M. (2013). How to escape from crop-to-weed gene flow: Phenological variation and isolation-by-time within weedy sunflower populations. *NEW PHYTOLOGIST*, *197*(2), 642–654. https://doi.org/10.1111/nph.12045

Rubio Teso, M. L., Álvarez Muñiz, C., Gaisberger, H., Kell, S., Lara-Romero, C., Magos Brehm, J., Maxted, N., & Iriondo, J. M. (2020). *In situ plant genetic resources in Europe: Crop wild relatives.* Farmer's Pride: Networking, partnerships and tools to enhance in situ conservation of European plant genetic resources. https://more.bham.ac.uk/farmerspride/wp-content/uploads/sites/19/2020/10/D1.2_In_situ_PGR_in_Europe_crop_wild_relatives.pdf

Ruhsam, M., Jessop, W., Cornille, A., Renny, J., & Worrell, R. (2019). Crop-to-wild introgression in the European wild apple Malus sylvestris in Northern Britain. *FORESTRY*, *92*(1), 85–96. https://doi.org/10.1093/forestry/cpy033

Saban, J. M., Romero, A. J., Ezard, T. H. G., & Chapman, M. A. (2023). Extensive crop-wild hybridization during Brassica evolution and selection during the domestication and diversification of Brassica crops. *Genetics*, *223*(4). Scopus. https://doi.org/10.1093/genetics/iyad027

Sanvido, O., Romeis, J., Gathmann, A., Gielkens, M., Raybould, A., & Bigler, F. (2012). Evaluating environmental risks of genetically modified crops: Ecological harm criteria for regulatory decision-making. *Environmental Science & Policy*, *15*(1), 82–91. https://doi.org/10.1016/j.envsci.2011.08.006

Scalabrin, S., Magris, G., Liva, M., Vitulo, N., Vidotto, M., Scaglione, D., Del Terra, L., Ruosi, M. R., Navarini, L., Pellegrino, G., Berny Mier y Teran, J. C., Toniutti, L., Suggi Liverani, F., Cerutti, M., Di Gaspero, G., & Morgante, M. (2024). A chromosome-scale assembly reveals chromosomal aberrations and exchanges generating genetic diversity in Coffea arabica germplasm. *Nature Communications*, *15*(1), 463. https://doi.org/10.1038/s41467-023-44449-8

Scarcelli, N., Chaïr, H., Causse, S., Vesta, R., Couvreur, T., & Vigouroux, Y. (2017). Crop wild relative conservation: Wild yams are not that wild. *BIOLOGICAL CONSERVATION*, *210*, 325–333. https://doi.org/10.1016/j.biocon.2017.05.001

Schierenbeck, K. A., & Ellstrand, N. C. (2009). Hybridization and the evolution of invasiveness in plants and other organisms. *Biological Invasions*, *11*(5), 1093–1105. https://doi.org/10.1007/s10530-008-9388-x

Schnitzler, A., Arnold, C., Cornille, A., Bachmann, O., & Schnitzler, C. (2014). Wild European Apple (Malus sylvestris (L.) Mill.) Population Dynamics: Insight from Genetics and Ecology in the Rhine Valley. Priorities for a Future Conservation Programme. *PLOS ONE*, *9*(5). https://doi.org/10.1371/journal.pone.0096596

Schreiber, M., Himmelbach, A., Börner, A., & Mascher, M. (2019). Genetic diversity and relationship between domesticated rye and its wild relatives as revealed through genotyping-by-sequencing. *Evolutionary Applications*, *12*(1), 66–77. Scopus. https://doi.org/10.1111/eva.12624

Setter, D., Mousset, S., Cheng, X., Nielsen, R., DeGiorgio, M., & Hermisson, J. (2020). VolcanoFinder: Genomic scans for adaptive introgression. *PLOS Genetics*, *16*(6), e1008867. https://doi.org/10.1371/journal.pgen.1008867

Shivrain, V., Burgos, N., Agrama, H., Lawton-Rauh, A., Lu, B., Sales, M., Boyett, V., Gealy, D., & Moldenhauer, K. (2010). Genetic diversity of weedy red rice (Oryza sativa) in Arkansas, USA. *WEED RESEARCH*, *50*(4), 289–302. https://doi.org/10.1111/j.1365-3180.2010.00780.x

Singer, M. F., & Berg, P. (1976). Recombinant DNA: NIH Guidelines. *Science*, *193*(4249), 186–188. https://doi.org/10.1126/science.11643320

Spear, M., Levi, S., Etterson, J., & Gross, B. (2023). Resurrecting urban sunflowers: Phenotypic and molecular changes between antecedent and modern populations separated by 36?years. *MOLECULAR ECOLOGY*, *32*(19), 5241–5259. https://doi.org/10.1111/mec.17112

Stevens, A., Hersi, M., Garritty, C., Hartling, L., Shea, B. J., Stewart, L. A., Welch, V. A., & Tricco, A. C. (2024). Rapid review method series: Interim guidance for the reporting of rapid reviews. *BMJ Evidence-Based Medicine*. https://doi.org/10.1136/bmjebm-2024-112899

Stewart, C. N., Halfhill, M. D., & Warwick, S. I. (2003). Transgene introgression from genetically modified crops to their wild relatives. *Nature Reviews Genetics*, *4*(10), 806–817. https://doi.org/10.1038/nrg1179

Stokstad, E. (2018, July 25). European court ruling raises hurdles for CRISPR crops. *Science*. https://www.science.org/content/article/european-court-ruling-raises-hurdles-crispr-crops

Stokstad, E. (2023, July 7). European Commission proposes loosening rules for gene-edited plants. *Science*. https://www.science.org/content/article/european-commission-proposes-loosening-rules-gene-edited-plants

Suarez-Gonzalez, A., Lexer, C., & Cronk, Q. C. B. (2018). Adaptive introgression: A plant perspective. *Biology Letters*, *14*(3), 20170688. https://doi.org/10.1098/rsbl.2017.0688

Tenaillon, M. I., Burban, E., Huynh, S., Wojcik, A., Thuillet, A.-C., Manicacci, D., Gérard, P. R., Alix, K., Belcram, H., Cornille, A., Brault, M., Stevens, R., Lagnel, J., Dogimont, C., Vigouroux, Y., & Glémin, S. (2023). Crop domestication as a step toward reproductive isolation. *American Journal of Botany*, *110*(7), e16173. https://doi.org/10.1002/ajb2.16173

Thomson, F. J., Letten, A. D., Tamme, R., Edwards, W., & Moles, A. T. (2018). Can dispersal investment explain why tall plant species achieve longer dispersal distances than short plant species? *New Phytologist*, *217*(1), 407–415. https://doi.org/10.1111/nph.14735

Todesco, M., Pascual, M. A., Owens, G. L., Ostevik, K. L., Moyers, B. T., Hübner, S., Heredia, S. M., Hahn, M. A., Caseys, C., Bock, D. G., & Rieseberg, L. H. (2016). Hybridization and extinction. *Evolutionary Applications*, *9*(7), 892–908. https://doi.org/10.1111/eva.12367

Tricou, T., Tannier, E., & de Vienne, D. M. (2022). Ghost Lineages Highly Influence the Interpretation of Introgression Tests. *Systematic Biology*, *71*(5), 1147–1158. https://doi.org/10.1093/sysbio/syac011

Trtikova, M., Lohn, A., Binimelis, R., Chapela, I., Oehen, B., Zemp, N., Widmer, A., & Hilbeck, A. (2017). Teosinte in Europe – Searching for the Origin of a Novel Weed. *Scientific Reports*, *7*(1), 1560. https://doi.org/10.1038/s41598-017-01478-w

Uwimana, B., D'Andrea, L., Felber, F., Hooftman, D., den Nijs, H., Smulders, M., Visser, R., & van de Wiel, C. (2012). A Bayesian analysis of gene flow from crops to their wild relatives: Cultivated (Lactuca sativa L.) and prickly lettuce (L. serriola L.) and the recent expansion of L. serriola in Europe. *MOLECULAR ECOLOGY*, *21*(11), 2640–2654. https://doi.org/10.1111/j.1365-294X.2012.05489.x

van Treuren, R., Hoekstra, R., & van Hintum, T. J. L. (2017). Inventory and prioritization for the conservation of crop wild relatives in The Netherlands under climate change. *Biological Conservation*, *216*, 123–139. https://doi.org/10.1016/j.biocon.2017.10.003

van Treuren, R., Hoekstra, R., Wehrens, R., & van Hintum, T. (2020). Effects of climate change on the distribution of crop wild relatives in the Netherlands in relation to conservation status and ecotope variation. *Global Ecology and Conservation*, *23*, e01054. https://doi.org/10.1016/j.gecco.2020.e01054

Veller, C., Edelman, N. B., Muralidhar, P., & Nowak, M. A. (2023). Recombination and selection against introgressed DNA. *Evolution*, *77*(4), 1131–1144. https://doi.org/10.1093/evolut/qpad021

Verleysen, L., Depecker, J., Bollen, R., Asimonyio, J., Hatangi, Y., Kambale, J.-L., Mwanga, I. M., Ebele, T., Dhed'a, B., Stoffelen, P., Ruttink, T., Vandelook, F., & Honnay, O. (2024). Crop-to-wild gene flow in wild coffee species: The case of Coffea canephora in the Democratic Republic of the Congo. *Annals of Botany*, *133*(7), 917–930. Scopus. https://doi.org/10.1093/aob/mcae034

Vi, T., Vigouroux, Y., Cubry, P., Marraccini, P., Phan, H. V., Khong, G. N., & Poncet, V. (2023). Genome-Wide Admixture Mapping Identifies Wild Ancestry-of-Origin Segments in Cultivated Robusta Coffee. *Genome Biology and Evolution*, *15*(5), evad065. https://doi.org/10.1093/gbe/evad065

Wang, H., Vieira, F. G., Crawford, J. E., Chu, C., & Nielsen, R. (2017). Asian wild rice is a hybrid swarm with extensive gene flow and feralization from domesticated rice. *Genome Research*, *27*(6), 1029–1038. https://doi.org/10.1101/gr.204800.116

Wang, T., van Dijk, A. D. J., Bucher, J., Liang, J., Wu, J., Bonnema, G., & Wang, X. (2023). Interploidy Introgression Shaped Adaptation during the Origin and Domestication History of Brassica napus. *Molecular Biology and Evolution*, *40*(9), msad199. https://doi.org/10.1093/molbev/msad199

Wang, X., Chen, L., & Ma, J. (2019). Genomic introgression through interspecific hybridization counteracts genetic bottleneck during soybean domestication. *Genome Biology*, *20*(1). Scopus. https://doi.org/10.1186/s13059-019-1631-5

Wedger, M., Roma-Burgos, N., & Olsen, K. (2022). Genomic revolution of US weedy rice in response to 21st century agricultural technologies. *COMMUNICATIONS BIOLOGY*, *5*(1). https://doi.org/10.1038/s42003-022-03803-0

Wen, D., Yu, Y., Zhu, J., & Nakhleh, L. (2018). Inferring Phylogenetic Networks Using PhyloNet. *Systematic Biology*, *67*(4), 735–740. https://doi.org/10.1093/SYSBIO/SYY015

Whitlock, M. C., & McCauley, D. E. (1999). Indirect measures of gene flow and migration: FST≠1/(4Nm+1). *Heredity*, *82*(2), 117–125. https://doi.org/10.1038/sj.hdy.6884960

Wilson, G. A., & Rannala, B. (2003). Bayesian Inference of Recent Migration Rates Using Multilocus Genotypes. *Genetics*, *163*(3), 1177–1191. https://doi.org/10.1093/genetics/163.3.1177

Wolf, D. E., Takebayashi, N., & Rieseberg, L. H. (2001). Predicting the Risk of Extinction through Hybridization. *Conservation Biology*, *15*(4), 1039–1053. https://doi.org/10.1046/j.1523-1739.2001.0150041039.x

Wu, D., Xie, L., Sun, Y., Huang, Y., Jia, L., Dong, C., Shen, E., Ye, C., Qian, Q., & Fan, L. (2023). A syntelog-based pan-genome provides insights into rice domestication and de-domestication. *GENOME BIOLOGY*, *24*(1). https://doi.org/10.1186/s13059-023-03017-5

Yakimowski, S. B., & Rieseberg, L. H. (2014). The role of homoploid hybridization in evolution: A century of studies synthesizing genetics and ecology. *American Journal of Botany*, *101*(8), 1247–1258. https://doi.org/10.3732/ajb.1400201

Zhang, Y., Zhu, Q., Shao, Y., Jiang, Y., Ouyang, Y., Zhang, L., & Zhang, W. (2023). Inferring Historical Introgression with Deep Learning. *SYSTEMATIC BIOLOGY*, *72*(5), 1013–1038. https://doi.org/10.1093/sysbio/syad033

Zunino, L., Cubry, P., Sarah, G., Mournet, P., El Bakkali, A., Aqbouch, L., Sidibé-Bocs, S., Costes, E., & Khadari, B. (2024). Genomic evidence of genuine wild versus admixed olive populations evolving in the same natural environments in western Mediterranean Basin. *PLOS ONE*, *19*(1). https://doi.org/10.1371/journal.pone.0295043

Annex A Priority taxa

A.1 Genera with CWR in the Netherlands or Europe and/or NGT plants

Table A.1.1. Target genera with crops and/or crop wild relatives (CWR) that satisfy one of three conditions: 1) CWR are identified as priority taxa for Europe (Rubio Teso et al., 2020), 2) CWR occur in the Netherlands (CWRnl.nl: van Treuren et al., 2017, 2020), or 3) crops have been identified as the subject of active R&D projects using New Genomic Techniques (Anses, 2024). Numbers indicate the number of (priority) CWR belonging to that genus in Europe (EU_CWR) or the Netherlands (NL_CWR), and the number of cultivated species for which R&D projects were reported, with the number of individual projects in brackets (RD_NGT). Genera with species satisfying all three conditions are highlighted in green; those with European priority CWR and crops in R&D using NGTs are highlighted in yellow; and those with crops in R&D using NGTs are highlighted in red. There were no genera with crops in R&D using NGTs and CWR in the Netherlands, without priority taxa in the EU. Genera with species reviewed in this report are indicated with an asterisk. Data from different sources was merged by matching genera to their most recent accepted name and stable taxonomic identifier (ID) according to the World Checklist of Vascular Plants (WCVP) and the International Plant Names index (IPNI).

Order	Family	Genus	EU_CWR	NL_CWR	RD_NGT	WCVP_ID	IPNI_ID
Apiales	Apiaceae	*Anthriscus*	0	3	0	2641791	331353-2
Apiales	Apiaceae	*Apium*	0	1	0	2643985	39683-1
Apiales	Apiaceae	*Carum*	0	1	0	2701447	39776-1
Apiales	Apiaceae	*Daucus **	3	1	1 (1)	2757895	39902-1
Apiales	Apiaceae	*Helosciadium*	0	3	0	2844878	40073-1
Apiales	Apiaceae	*Pastinaca*	0	1	0	2403460	40341-1
Apiales	Apiaceae	*Pimpinella*	0	2	0	2402470	30043322-2
Apiales	Apiaceae	*Sison*	0	1	0	2473818	40539-1
Apiales	Apiaceae	*Trocdaris*	0	1	0	2438034	40645-1
Aquifoliales	Aquifoliaceae	*Ilex*	0	1	0	2860299	60437220-2
Arecales	Arecaceae	*Phoenix*	3	0	0	152643	31428-1
Asparagales	Amaryllidaceae	*Allium*	20	6	0	294888	30000901-2
Asparagales	Asparagaceae	*Asparagus*	16	2	0	274909	30275681-2
Asterales	Asteraceae	*Astartoseris*	1	0	0	3135030	77167416-1
Asterales	Asteraceae	*Carthamus*	8	0	0	3134274	8120-1
Asterales	Asteraceae	*Cichorium*	5	1	0	3134297	30096161-2
Asterales	Asteraceae	*Cynara*	6	0	0	3134409	8514-1
Asterales	Asteraceae	*Helianthus **	7	0	0	3133724	30000729-2
Asterales	Asteraceae	*Lactuca **	11	2	1 (1)	3134294	327393-2
Asterales	Asteraceae	*Scorzonera*	0	1	0	3134279	30092856-2
Brassicales	Brassicaceae	*Arabidopsis*	0	1	0	2645200	12326-1
Brassicales	Brassicaceae	*Armoracia*	1	1	0	2652936	331369-2
Brassicales	Brassicaceae	*Barbarea*	1	3	0	2669601	331385-2
Brassicales	Brassicaceae	*Brassica **	19	4	1 (3)	2681967	30043511-2

Order	Family	Genus	EU_CWR	NL_CWR	RD_NGT	WCVP_ID	IPNI_ID
Brassicales	Brassicaceae	*Camelina*	0	0	1 (4)	2694356	12396-1
Brassicales	Brassicaceae	*Capsella*	0	1	0	2697963	331460-2
Brassicales	Brassicaceae	*Coincya*	1	1	0	2730462	12460-1
Brassicales	Brassicaceae	*Crambe*	15	1	0	2741235	12477-1
Brassicales	Brassicaceae	*Descurainia*	0	1	0	2761690	331548-2
Brassicales	Brassicaceae	*Diplotaxis*	5	2	0	2771218	12525-1
Brassicales	Brassicaceae	*Eruca*	1	0	0	2798706	12574-1
Brassicales	Brassicaceae	*Erucastrum*	2	1	0	2798808	12578-1
Brassicales	Brassicaceae	*Isatis*	0	1	0	2866871	30287110-2
Brassicales	Brassicaceae	*Lepidium*	2	7	0	2338293	60437330-2
Brassicales	Brassicaceae	*Moricandia*	1	0	0	2360751	12813-1
Brassicales	Brassicaceae	*Raphanus*	2	1	0	2422455	12986-1
Brassicales	Brassicaceae	*Rorippa*	2	3	0	2417180	30012174-2
Brassicales	Brassicaceae	*Sinapidendron*	5	0	0	2476363	13042-1
Brassicales	Brassicaceae	*Sinapis*	3	1	0	2476410	13043-1
Brassicales	Brassicaceae	*Thlaspi*	0	0	1 (1)	2437720	13125-1
Caryophyllales	Amaranthaceae	*Atriplex*	1	8	0	2664993	30000935-2
Caryophyllales	Amaranthaceae	*Beta*	8	1	1 (1)	2675545	30253697-2
Caryophyllales	Amaranthaceae	*Blitum*	0	2	0	2677584	329881-2
Caryophyllales	Amaranthaceae	*Chenopodiastrum*	0	2	0	2913108	77121006-1
Caryophyllales	Amaranthaceae	*Chenopodium*	4	3	0	2716928	60437273-2
Caryophyllales	Amaranthaceae	*Lipandra*	0	1	0	2339159	6915-1
Caryophyllales	Amaranthaceae	*Oxybasis*	0	3	0	2392552	6952-1
Caryophyllales	Amaranthaceae	*Patellifolia*	1	0	0	2409105	6957-1
Caryophyllales	Amaranthaceae	*Salsola*	1	2	0	2480904	30012872-2
Cucurbitales	Cucurbitaceae	*Citrullus*	2	0	0	2723887	331494-2
Cucurbitales	Cucurbitaceae	*Cucumis*	2	0	1 (2)	2746873	30023062-2
Dipsacales	Caprifoliaceae	*Valerianella*	0	4	0	2463556	30019025-2
Ericales	Ericaceae	*Vaccinium*	3	5	0	2459130	30000401-2
Fabales	Fabaceae	*Astragalus*	3	1	0	2657136	330028-2
Fabales	Fabaceae	*Cicer*	6	0	1 (1)	2720544	22016-1
Fabales	Fabaceae	*Coronilla*	1	1	0	2737916	22094-1
Fabales	Fabaceae	*Cytisus*	0	1	0	2754521	60455196-2
Fabales	Fabaceae	*Galega*	1	0	0	2816216	328382-2
Fabales	Fabaceae	*Glycine* *	0	0	1 (4)	3223627	331628-2
Fabales	Fabaceae	*Hedysarum*	1	0	0	2840943	30007304-2

Order	Family	Genus	EU_CWR	NL_CWR	RD_NGT	WCVP_ID	IPNI_ID
Fabales	Fabaceae	*Lathyrus*	25	8	0	2352057	30005985-2
Fabales	Fabaceae	*Lotus*	3	3	1 (1)	2344629	30016283-2
Fabales	Fabaceae	*Lupinus*	7	1	0	2348057	60436956-2
Fabales	Fabaceae	*Medicago*	28	6	2 (5)	2368626	331761-2
Fabales	Fabaceae	*Melilotus*	2	3	0	2367483	325467-2
Fabales	Fabaceae	*Onobrychis*	1	1	0	2389110	30311989-2
Fabales	Fabaceae	*Ornithopus*	2	1	0	2387919	23086-1
Fabales	Fabaceae	*Trifolium*	18	13	0	2440723	325471-2
Fabales	Fabaceae	*Trigonella*	1	0	0	2439377	30107635-2
Fabales	Fabaceae	*Vicia*	32	10	0	2452522	330034-2
Fabales	Fabaceae	*Vigna*	0	0	1 (1)	2451479	325971-2
Fagales	Betulaceae	*Corylus **	3	1	0	47827	13489-1
Fagales	Fagaceae	*Castanea*	2	1	0	34916	30004045-2
Fagales	Juglandaceae	*Juglans*	5	0	0	2331809	30000723-2
Lamiales	Lamiaceae	*Mentha **	1	5	0	124319	30016176-2
Lamiales	Lamiaceae	*Origanum*	0	1	0	143731	21080-1
Lamiales	Lamiaceae	*Salvia*	0	0	1 (1)	181875	30000096-2
Lamiales	Lamiaceae	*Thymus*	0	3	0	204357	30002942-2
Lamiales	Oleaceae	*Olea **	1	0	0	355240	328104-2
Lamiales	Plantaginaceae	*Plantago*	1	0	0	2556648	30001135-2
Malpighiales	Linaceae	*Linum*	5	1	0	2347347	24945-1
Malpighiales	Salicaceae	*Populus*	0	0	1 (1)	2917745	328417-2
Malvales	Malvaceae	*Gossypium*	0	0	1 (1)	2830979	30013018-2
Malvales	Malvaceae	*Theobroma*	0	0	1 (1)	2519778	30005713-2
Myrtales	Myrtaceae	*Myrtus*	1	0	0	132268	30000914-2
Poales	Poaceae	*Aegilops **	20	0	0	450699	17369-1
Poales	Poaceae	*Agropyron*	5	0	0	451053	327635-2
Poales	Poaceae	*Agrostis*	3	4	0	451057	325863-2
Poales	Poaceae	*Alopecurus*	1	5	0	451067	17424-1
Poales	Poaceae	*Arrhenatherum*	1	1	0	451084	30158211-2
Poales	Poaceae	*Avena*	11	1	1 (1)	451095	17560-1
Poales	Poaceae	*Bromus*	0	3	0	451165	30000273-2
Poales	Poaceae	*Cynodon*	1	1	0	451328	331181-2
Poales	Poaceae	*Dactylis*	1	1	0	451339	30077575-2
Poales	Poaceae	*Digitaria*	0	2	0	451392	329395-2
Poales	Poaceae	*Echinochloa*	2	1	0	451441	17976-1

Order	Family	Genus	EU_CWR	NL_CWR	RD_NGT	WCVP_ID	IPNI_ID
Poales	Poaceae	*Eleusine*	2	0	0	451451	30007287-2
Poales	Poaceae	*Elymus*	9	2	0	451462	328322-2
Poales	Poaceae	*Eragrostis*	2	0	0	451476	30000628-2
Poales	Poaceae	*Festuca*	3	4	0	451511	328907-2
Poales	Poaceae	*Hordeum* *	5	3	1 (2)	419181	327148-2
Poales	Poaceae	*Leymus*	4	1	0	422547	328321-2
Poales	Poaceae	*Lolium*	7	5	1 (1)	422752	18423-1
Poales	Poaceae	*Oryza* *	1	0	1 (44)	426595	18639-1
Poales	Poaceae	*Panicum*	1	0	1 (2)	427062	1194715-2
Poales	Poaceae	*Phalaris*	3	1	0	433365	18717-1
Poales	Poaceae	*Phleum*	2	3	0	433709	18725-1
Poales	Poaceae	*Poa*	5	9	0	434980	30001404-2
Poales	Poaceae	*Saccharum*	1	0	0	439841	330004-2
Poales	Poaceae	*Secale* *	4	0	1 (2)	441721	18991-1
Poales	Poaceae	*Setaria*	1	3	0	442152	331287-2
Poales	Poaceae	*Sorghum* *	2	0	0	443224	331290-2
Poales	Poaceae	*Trisetum*	1	1	0	447705	19214-1
Poales	Poaceae	*Triticum* *	3	0	2 (8)	448335	30034910-2
Poales	Poaceae	*Zea* *	0	0	1 (12)	450340	19280-1
Ranunculales	Papaveraceae	*Papaver*	1	3	1 (1)	2560992	30000445-2
Rosales	Cannabaceae	*Humulus*	1	1	0	2855034	30027247-2
Rosales	Moraceae	*Ficus*	1	0	0	2809363	327905-2
Rosales	Rosaceae	*Comarum*	1	1	0	2943883	30063938-2
Rosales	Rosaceae	*Fragaria* *	5	2	1 (1)	2948599	30014957-2
Rosales	Rosaceae	*Malus*	3	1	0	2950387	30024474-2
Rosales	Rosaceae	*Prunus*	25	3	0	2958711	30003057-2
Rosales	Rosaceae	*Pyrus*	9	1	0	2962753	30000967-2
Rosales	Rosaceae	*Rubus*	8	5	0	2976315	30000199-2
Sapindales	Anacardiaceae	*Pistacia*	5	0	0	2408144	1883-1
Sapindales	Rutaceae	*Citrus*	0	0	2 (2)	2723938	30022289-2
Saxifragales	Grossulariaceae	*Ribes*	9	4	0	2426003	30002461-2
Solanales	Solanaceae	*Nicotiana*	0	0	1 (1)	2383073	325974-2
Solanales	Solanaceae	*Solanum* *	5	2	2 (18)	2936484	30000630-2
Vitales	Vitaceae	*Vitis*	7	0	1 (1)	2456523	325876-2
Zingiberales	Musaceae	*Musa*	0	0	1 (1)	254737	327926-2
Total			493	222	35 (126)		

A.2 Traits introduced in NGT plants

Table A.2.1. Traits introduced in NGT plants, from the early (light blue) and advanced (blue) R&D stage to the commercial stage (dark blue). Data was retrieved from the French National Health and Safety Agency of Food, Environment and Labour (Anses, 2024). Trait categories were originally described in (Parisi & Rodríguez-Cerezo, 2021). Species reviewed in this report are indicated with an asterisk.

Vernacular name	Scientific name	Storage performance	Yield and architecture	Modified colour/flavour	Modified composition	Breeding tools	Biotic stress tolerance	Abiotic stress tolerance	Herbicide tolerance	Other traits	Total
Oat	*Avena sativa*						1				1
Sugar beet	*Beta vulgaris*			1			1				2
Rapeseed	*Brassica napus* *		2		1						3
Camelina	*Camelina sativa*				4						4
Chickpea	*Cicer arietinum*							1			1
Grapefruit	*Citrus paradisi*						1				1
Orange	*Citrus sinensis*						1				1
Cucumber	*Cucumis sativus*					1	1				2
Carrot	*Daucus carota* *				1						1
Wild strawberry	*Fragaria vesca* *		1								1
Soybean	*Glycine max* *							1	1	2	4
Cotton	*Gossypium hirsutum*		1								1
Barley	*Hordeum vulgare* *		1			1					2
Lettuce	*Lactuca sativa* *							1			1
Ryegrass	*Lolium perenne*				1						1
Trefoil	*Lotus japonicus*		1								1
Alfalfa	*Medicago sativa*									2	2
Barrelclover	*Medicago truncatula*				1					2	3
Banana	*Musa spp.*						1				1
Tobacco	*Nicotiana benthamiana*										0
Rice	*Oryza sativa* *		15	3	7	8	5	2	4		44
Switchgrass	*Panicum virgatum*		1		1						2
Opium poppy	*Papaver somniferum*				1						1
Poplar	*Populus sp.*				1						1
Sage	*Salvia miltiorrhiza*				1						1
Rye	*Secale cereale* *					1	1				2
Tomato	*Solanum lycopersicum*	1	2	4	4		3			1	15
Potato	*Solanum tuberosum*				2				1		3
Cocoa	*Theobroma cacao*						1				1
Field pennycress	*Thlaspi arvense*				1						1
Common wheat	*Triticum aestivum* *		2		1	1	2	1			7
Durum wheat	*Triticum durum*				1						1
Cowpea	*Vigna unguiculata*		1								1
Grape vine	*Vitis vinifera*						1				1
Maize	*Zea mays* *				3	5		3		1	12
Total		1	27	8	31	17	19	9	6	8	126

Annex B Surveyed literature

B.1 Summary per species

Below we give and overview of the species reviewed and their main findings, organised in alphabetical order by their botanical family. Names of crops with crop wild relatives in the Netherlands are indicated in bold; those whose relatives are priority taxa in the EU are underscored.

Apiaceae (Apiales)

Carrot: Hernández et al. (2023) studied 35 thousand SNPs generated by GBS in almost 600 individuals of cultivated and wild carrot (*Daucus carota*) across the United States. With a combination of global ancestry analysis (using ADMIXTURE) and D-statistics, they showed that 24% of the wild carrot population contained >10% cultivated ancestry, and that in 26% of ABBA-BABA tests *D* deviated significantly from zero. Use of the *f4*-ratio revealed that 15-24% of the genome was introgressed. Crop alleles were not found in all areas where wild and cultivated populations are sympatric, suggesting that in some places reproductive barriers may have evolved, or that there may be selection against them.

Asteraceae (Asterales)

Lettuce: Uwimana et al. (2012) used ten microsatellites to genotype thousands of cultivated lettuce (*Lactuca sativa*) and its weedy relative prickly lettuce (*Lactuca serriola*). Through global ancestry analysis with STRUCTURE, they found that 7% of the weedy population contained >10% cultivated ancestry. They also used the NewHybrids software to classify different generations of hybrids, which uncovered most of the admixed individuals identified by STRUCTURE, including some additional advanced backcrosses. A putative motor behind the northward expansion of prickly lettuce in Europe, crop-to-wild gene flow was found to be the least common in Northern Europe, suggesting that the introgression of local crop alleles was not the key to this range shift and that expansion may have been driven by ecological factors instead.

Sunflower: Originating from North America, sunflower introgression into its wild and weedy relatives has been studied on multiple continents, including North and South America and in Europe. In all these locations gene flow between cultivated and wild sunflowers is facilitated by overlapping flowering times (although in France some divergence is observed), shared pollinators, and the outcrossing nature of the species.

South America: In Argentina, Gutierrez et al. (2010) used 51 random amplified polymorphic DNA (RAPD) markers to genotype individuals from 9 cultivated sunflower populations (*Helianthus annuus*) and 26 populations of its weedy relative *Helianthus petiolaris*. By identifying crop diagnostic alleles, they were able to detect that 0-2% wild individuals were introgressed in each population, with an average of 0.3%. Mondon et al. (2018) confirmed hybridisation between these species in Argentina with ABBA-BABA tests, applied to 3526 SNPs generated by GBS for 182 individuals. Gene flow was tested for multiple quartets including Argentinian and North American wild *H. annuus*, and show consistently positive D for almost all population pairs tested.

North America: Another study in North America used the *fd* statistic to show that introgression from cultivated sunflower accounted for 7.1% and 9.7% of the genomes of two of its wild relatives (*H. argophyllus* and *bolanderi-exilis*, respectively), using 4645 SNPs generated by GBS for almost three hundred individuals (Baute et al., 2016). Spear et al. (2023) used 3427 SNPs generated with RNA-seq data from 39 individuals germinated in a resurrection experiment spanning 36 years of urban sunflower evolution in Minneapolis (United States). Through a global ancestry analysis with STRUCTURE, they showed that over the span of several decades, the proportion of wild *H. annuus* individuals with ancestry from domesticated sunflowers rose from 0% to 7.1% in central Minneapolis and 13.6% in northern Minneapolis. Coupled with temporal shifts in gene expression and phenotypes, the authors speculate that the observed increase in crop-to-wild introgression could have played a role in adaptation to climate change.

Europe: In France and Spain, *H. annuus* has emerged as an agricultural weed. Muller et al. (2011) genotyped more than 300 individuals with 16 microsatellites and performed a STRUCTURE analysis showing the mixed ancestry of the weedy population (although exact proportions in the population or genomic extent were not reported). The detection of a crop-specific mitochondrial marker in all weedy individuals shows that crop-to-wild introgression happened through cultivated sunflower pollination by wild sunflower. A follow-up study by Roumet et al. (2013) used the same 16 microsatellites to genotype more than 700 individuals, and performed parentage analysis to determine that 6.7-38.2% of weedy individuals had a recently cultivated parent. A positive correlation between genetic divergence and flowering time of the weedy type suggests partial reproductive isolation due to phenological differences.

Betulaceae (Fagales)

Hazelnut: Helmstetter et al. (2020) applied double-digest restriction-site associated DNA (ddRAD) to generate 60 thousand SNPs across the genomes of wild and cultivated European hazelnut (*Corylus avellana*) in Turkey, which are conspecific and have comparable levels of genetic diversity. Using a combination of *D*-statistics and population graph inference with TreeMix, they inferred three gene flow events from crop to crop, wild to crop and crop to wild populations, of which the latter was supported by an ABBA-BABA test. The obligate outcrossing nature of hazelnut and the fact that wild and cultivated hazelnut are conspecific may explain why admixture between different cultivars and wild hazelnut has been recurrent.

Brassicaceae (Brassicales)

Saban et al. (2023) analysed more than 5 million SNPs in 79 representatives of *Brassica* species to disentangle the relationships and patterns of gene flow during the evolution of various crops in this genus. Their broad taxonomic sampling allowed a comparison of introgression signals between lineages, disentangling ancient from more recent gene flow.

Cabbage: A PhyloNet analysis supported hybrid origins of multiple *B. oleracea* varieties (subsp. *alboglabra* and subsp. *botrytis*), and one wild *B. oleracea* population. D-statistics further supported gene flow between wild and domesticated *B. oleracea* and almost all other monophyletic CWR (*B. cretica, B. insularis, B. macrocarpa* and *B. rupestris*). The genomic extent of introgression was investigated with *fd*. While no signal of introgression from domesticated *B. oleracea* into the other CWR was found, introgression from wild and cultivated *B. oleracea* into *B. cretica* accounted for 9.46–14.28% of the genome (Saban et al., 2023).

Turnip: A PhyloNet analysis supported hybrid origins of *B. rapa* subsp. *trilocuralis*. The *D*-statistic provided evidence of introgression between multiple *B. rapa* varieties and monophyletic CWR. The extent was quantified with *fd*. In addition to small amounts of wild-to-crop introgression, crop-to-wild introgression was only found for, *B. rupestris*, and not for *B. cretica* or *B. macrocarpa*, amounting to 0.22–0.88% of the genome (Saban et al., 2023).

Rapeseed: Using whole-genome sequencing Wang et al. (2023) found evidence for extensive interploid gene flow during the domestication history of different crop-types of *Brassica napus*, which emerged by allopolyploidisation between *B. rapa* a wild *B. oleracea* and contains introgression footprints from diploid relatives of both its ancestors. While the study focused on the role of crop-to-crop and wild-to-crop introgression during crop evolution, both the *D*-statistic and *f4*-ratio demonstrated gene flow between wild *B. oleracea* and multiple lineages of cultivated *B. napus*, including Siberian kale, winter rapeseed and spring rapeseed; although these analyses didn't reveal the direction of gene flow. A TreeMix analysis identified numerous migration events between several crop lineages, with the only crop-to-wild migration edge inferred from curly kale (*B. oleracea*) into wild *B. oleracea*.

Fabaceae (Fabales)

Lima bean: Heredia-Pech et al. (2022) analysed 83 accessions of wild and cultivated lima bean (*Phaseolus lunatus*) from the Yucatan peninsula in Mexico with 15 thousand SNPs generated by GBS. Using a wide array of complementary analyses, including global ancestry inference with STRUCTURE, haplotype analysis using NGSEP, *D*- and *f*-statistics with Dsuite and migration rate with MIGRATE-N, they found that 36% of wild accessions had high levels of cultivated ancestry, amounting to varying proportions of the genome depending on whether they used *Df* (12-18%) or *f4* (15-24%). Gene-flow was found to be bi-directional, but asymmetric, with a higher rate of crop-to-wild gene flow in one region and a higher rate of wild-to-crop gene flow in another, possibly explained by different agricultural practices. The overall migration rate was higher from crops into admixed wild populations (M=157) than pure wild populations (M=114), with the first possibly acting as a bridge for the second.

Soy bean: Wang et al. (2019) used genome-wide SNPs generated from whole-genome sequencing (WGS) to study introgression between the soy bean (*Glycine max*) and its wild progenitor (*Glycine soja*). By counting the haplotypes that were identical by descent (IBD) in the different subpopulations, they quantified the relative IBD (rIBD) between wild and cultivated tracts in genomic windows, and identified outlier regions suggestive of introgression. This led to the identification of introgression tracts of cultivated haplotypes covering 0.059-41% of the wild genome, averaging 1.9% per accession. These regions also had a significantly lower value of the *D*-statistic.

Lamiaceae (Lamiales)

Spearmint: Olofsson et al. (2024) used whole genome sequencing to characterise the diversity of 93 herbarium specimens of the allotetraploid cultivated spearmint (*Mentha spicata*) and its presumed diploid parent species *M. suaveolens* and *M. longifolia* in the Nordics. With a PCA and a global ancestry analysis conducted with NGSadmix on more than a million SNPs, they found that cultivated mint is more closely related to *M. longifolia* than to *M. suaveolens*. While they found no cultivated ancestry in the wild progenitor species, *D*-statistics supported gene flow between *M. spicata* and *M. longifolia*, with positive signals found for 58% of the tested trios, varying from 53% (for the cultivated cluster that is intermediate between the two parents) to

63% (for the cluster that is more closely related to *M. longifolia*). These results demonstrate the interfertility of cultivated spearmint with at least one of its parents, despite differences in ploidy.

Oleaceae (Lamiales)

Olive: Zunino et al. (2024) analysed over 500 accessions of wild and cultivated olive (*Olea europaea*) in the western Mediterranean basin using target capture to generate over 140 thousand SNPs. Despite a presumed origin in the eastern Mediterranean, cultivated olive showed low genetic differentiation (F_{ST}) with a large number of western wild olive populations. Through analysis of global ancestry components (inferred by sNMF), the authors showed that many individuals in Morocco, Spain and France deviate from the dominant "pure wild" ancestry in the western Mediterranean, with signals of admixture from locally cultivated olives. Using TreeMix they inferred two migration events: one from cultivated French olive to wild populations in Morocco, France and Spain, and one from wild to cultivated or already admixed populations in Spain, with a migration weight of w=0.425 and w=0.485, respectively.

Poaceae (Poales)

Barley: Using a mixture of genotype-by-sequencing (GBS) and whole genome re-sequencing, Guo et al. (2022) studied the diversity and origins of barley (*Hordeum vulgare*) crop wild relatives *H. spontaneum* and *H. agriocrithon*[4]. They showed the latter to have arisen through hybridisation between cultivated barley and true wild barley (*H. spontaneum*), despite often presumed to be the wild progenitor. Global ancestry analysis with ADMIXTURE demonstrated the admixed nature of almost the entire *H. agriocrithon* population, and positive *f4*-ratios confirmed the occurrence of gene flow between the cultivated and wild lineages. Another study by Jin et al. (2024), which did not include *H. spontaneum or H. agriocrithon* but considered over thirty other wild relatives, used 103 gene alignments generated by target capture to explore relationships within the genus. While the authors did not implicate *H. vulgare* in any gene flow events, they detected numerous gene flow events among New World diploid species (which all have the same subgenome) by analysing discordant gene trees with QuIBL, supported by 2-8% of genomic windows. Additional ABBA-BABA tests, however, did not find any evidence for gene flow between the different subgenomes, which may reflect long-standing geographic separation.

Maize: Maize (*Zea mays* subsp. *mays*) wild relatives are collectively called "teosinte", and consist of multiple species and subspecies, including the wild progenitor of cultivated maize (*Zea mays* subp. *parviglumis*) and its close relative (*Zea mays* subsp. *mexicana*). Hufford et al. (2013) studied 189 individuals of nine sympatric maize-*mexicana* populations by sequencing 40 thousand SNPs and found asymmetric gene flow between maize landraces and *mexicana*, with older and stronger signals of wild-to-crop introgression than crop-to-wild introgression. HAPMIX and STRUCTURE identified 11.4% and 9.2% of the genome as opposite ancestry, respectively, but only 3.2% of the *mexicana* genome carried signals of introgression in both scans; *f3*-statistics were able to confirm gene flow between maize and *mexicana* in 8 out of 9 populations. Rojas-Barrera et al. (2019) studied the genetic contribution of modern maize varieties to traditional landraces and teosinte in Mexico by analysing more than 300 thousand

---

[4] Plants of the World Online (POWO) considers both of these species names synonyms of the same accepted taxon H. vulgare subsp. spontaneum. The authors of this study consider them separate taxa.

SNPs generated by GBS for 385 individuals, collected over a time span of 70 years. They used the *D*-statistic to test for evidence of introgression between sympatric and allopatric populations and found significant gene flow between modern varieties and sympatric *mexicana* across time. A stronger signal in more recently sampled teosinte indicates that gene flow is either intensifying or accumulating, although the genomic extent wasn't quantified. Outside Mexico, introgression from maize into teosinte has been found in France and Spain by Le Corre et al. (2020), who analysed 70 local teosinte with 25 thousand SNPs. Using *f4*-statistics, they found an even larger extent of maize introgression into *mexicana* than in Mexico, ranging from 12.2-42.2% and including herbicide resistance and flowering time genes. This was supported by a TreeMix analysis showing two gene flow events from the Dent inbred maize lines into French and Spanish teosinte with a migration weight of 0.14 and 0.39, respectively.

Pearl millet: Olodo et al. (2020) used 12 microsatellites to characterise genetic diversity and relationships among over a thousand accessions of cultivated and wild pearl millet (*Cenchrus americanus*) in Senegal. They analysed global ancestry with STRUCTURE to infer admixture proportions, and found that average crop-to-wild ancestry (0.033) was significantly higher than the average wild-to-crop ancestry (0.019). Pearl millet is an outcrossing crop known to produce hybrids, and since the authors deliberately focused on "pure" wild pearl millet and excluded weedy types, the observed extent of admixture is likely an underestimation. Their temporal dataset over a time period of 40 years allowed them to demonstrate that the gene pool is not only homogenizing due to gradual abandonment of the crop; the proportion of wild ancestry in cultivated pearl millet is also decreasing, which might be explained by contraction of the cultivation area and reduced overlap with the distribution of its wild progenitor.

Rice: The large number of studies on Asian rice (*Oryza sativa*) were roughly divided into two sets: those studying patterns of admixture in the evolution of weedy rice, and those studying introgression in the context of the multiple origins of rice, with the cultivated subspecies (*japonica*, *indica* and *aus*) each presumably domesticated from a different wild progenitor, but with potential gene flow from the earlier domesticate to later domesticates.

Weedy rice: Shivrain et al. (2010) used SSR (microsatellite) markers to run a STRUCTURE analysis on more than a hundred individuals of cultivated *japonica* and weedy *indica* rice in Arkansas, and identified that 25% of red rice accessions had high (>20%) portions of cultivated ancestry. Wedger et al. (2022) used genome-wide SNPs to infer both global (with ADMIXTURE) and local (with Loter) ancestry of 48 weedy rice accessions from Arkansas, and found that 73% of accessions had high (>15%) portions of cultivated *japonica* ancestry, with 26-31% of the genome traceable to cultivated rice, revealing adaptive introgression of cultivated alleles at the herbicide resistance gene *ALS*. Presotto et al. (2024) generated 15 thousand SNPs with GBS for 10 cultivated and 88 weedy rice accessions in Argentina, and analysed these with *D*- and *f4*-statistics to demonstrate admixture between the *aus*-type weeds and *indica*-type cultivars. While lack of genome-wide coverage precluded identifying the genomic regions involved, the admixed weeds showed improved herbicide resistance compared to the non-admixed weeds. A global analysis of genome-wide data of more than 500 weedy rice samples showed that, despite the feral origins of most weedy rice, hybridisation was especially common among Latin American weedy rice types, with 54% of samples showing mixed ancestry between *aus* and *indica* based on a fastSTRUCTURE analysis (Qiu et al., 2020). This finding was supported by a migration event from Latin American *aus* into Latin American mixed genotypes (sister to Latin American *indica*) inferred by a TreeMix analysis. While the percentages in these studies vary, one commonality is that introgression was exclusively detected between different subspecies, and that weedy populations pick up adaptive alleles from local cultivars during range expansion, partially driven by agriculturally determined selection pressures.

Domestication syndrome: Choi et al. (2017) used *D*-statistics on 19 thousand gene alignments (from whole-genome sequencing data) to provide evidence for gene flow between the *japonica*, *indica*, and *aus* subspecies of cultivated rice. With the Bayesian modelling program G-PhoCS, they subsequently quantified the level of gene flow from *japonica* to wild *O. nivara* (the progenitor species of domesticated *indica* and *aus*), or proto-domesticated forms of *indica* and *aus*, at 17% and 15%, respectively. The shared crop alleles between these three subspecies gave rise to the hypothesis that crop-to-wild gene flow played a role in the acquisition of a single domestication syndrome by separate lineages of wild rice. This "single domestication, multiple origins" hypothesis was later confirmed through analysis of 175 thousand syntelog groups in a pangenome of rice by Wu et al. (2023), who identified 57 introgression blocks, totalling 21.2 Mb in size and including most well-known domestication genes, using the *fd*-statistic.

<u>Rye</u>: Schreiber et al. (2019) used GBS to generate 55 thousand SNPs, with which they analysed the genetic diversity and relationships between hundreds of accessions of rye (*Secale cereale*) and its wild progenitor (*S. vavilovii*). Despite having sampled only five accessions of *S. vavilovii*, the authors found a significant excess of allele sharing between Armenian wild and domesticated rye using *D*-statistics. The relatively recent domestication history of rye (which may have had multiple independent origins, starting as an agricultural weed) and its outcrossing nature are both consistent with the high amounts of shared ancestry and low differentiation between the cultivated and wild species, making it likely that gene flow with its closest relatives is more widespread than what is observed with this limited sampling. Building on these results, Rabanus-Wallace et al. (2021) analysed 72 thousand SNPs generated with GBS for 950 samples, and found excess allele sharing among 6 out of 7 genetic clusters of rye supported by significant ABBA-BABA tests, including *S. strictum* and *S. sylvestre* in addition to *S. cereale* and *S. vavilovii*, further demonstrating the reticulate history of this genus and the incomplete reproductive isolation of rye from its wild relatives.

<u>Sorghum</u>: Two studies we surveyed confirm that crop-to-wild gene flow happens between cultivated sorghum (*Sorghum bicolor* subsp. *bicolor*) and its wild progenitor *Sorghum bicolor* subsp. *verticilliflorum.* Both used microsatellite markers to estimate the relationships and ancestry among ~500 accessions, identifying putatively admixed individuals with STRUCTURE. Mutegi et al. (2012) used the ancestry coefficients to determine that up to 28% of the genome of wild sorghum in Kenya was of cultivated origin, and identified asymmetric gene flow between cultivated and wild sorghum with BayesAss (with a crop-to-wild migration rate of 0.08 as compared to a wild-to-crop migration rate of 0.003). This asymmetry could be explained by differences in population size (with the cultivated populations being more abundant), mating system (with cultivated sorghum having a higher selfing rate than the wild) and agricultural practices (due to farmers recognising and selecting against wild-to-crop hybrids). While Adugna & Bekele (2017) did not give exact ancestry proportions, they also observed mixed ancestry between cultivated and wild sorghum and Ethiopia, and further calculated the migration rate between these populations based on F_{ST}, estimating crop-to-wild gene flow at $N_m = 2.13$ migrants per generation.

Wheat: Multiple studies characterised the patterns of gene flow between common wheat (*Triticum aestivum*) and its wild relatives in the goatgrass genus (*Aegilops*) in Spain and the southern Mediterranean area, each genotyping hundreds of individuals. Despite differences in ploidy (wheat is hexaploid, whereas *Aegilops* species have variable ploidy and can be tetraploid) and the selfing nature of the species, hybridisation between these genera appears to be common. AFLP studies have found high proportions of admixed individuals particularly *Ae. neglecta* and *Ae. triuncialis*, ranging from 10-25%, whereas wheat ancestry in *Ae. geniculata* was negligible. These ancestry coefficients were estimated with STRUCTURE (Arrigo et al., 2011) and InStruct (Pajkovic et al., 2014), and are consistent with a study of Parisod et al. (2013), who found 17% of an *Ae. triuncialis* population to carry wheat diagnostic alleles using microsatellites. In addition, Pajkovic et al. (2014) used Approximate Bayesian Computation (ABC) to infer that the current level of admixture in *Ae. triuncialis* could have emerged in less than 20 generations, despite a low crop-to-wild migration rate of 0.027% of the carrying capacity.

Rosaceae (Rosales)

Almond: Delplancke et al. (2012) used 12 microsatellite markers to study patterns of diversity, structure and gene flow among more than 500 samples of cultivated almond (*Prunus dulcis*) and its wild relative (*Prunus orientalis*) in the Eastern Mediterranean basin. Continuous gene flow was tested with the MIGRATE-N model, with which the authors found substantial symmetric gene flow in both directions (mutation scaled migration rates of 15.64 from crop-to-wild, and 16.96 from wild-to-crop), amounting to 28 migrants per generation. Both species were estimated to have roughly equal effective population sizes and levels of genetic diversity.

Apple: In total 11 of the studies we surveyed focused on apple, some of which are described in more detail as a case study (Box 2). These studies have all used one to several dozen microsatellites on hundreds to over a thousand samples of cultivated and one or more wild species of apple, covering much of their native distribution in Eurasia, and mostly inferred admixture from global ancestry analysis with the STRUCTURE model. Extensive crop-to-wild gene flow was found from domesticated apple (*Malus domestica*) into the wild progenitor of apple (*Malus sieversii*) in its centre of origin in Central Asia (Cornille et al., 2013; Feurtey et al., 2020; Ha et al., 2021; Omasheva et al., 2017). Bidirectional gene flow was also detected between domesticated apple and two other wild relatives (*Malus orientalis* and *Malus sylvestris*) in the Caucasus (Bina et al., 2022; Cornille et al., 2013) and in Europe (Chen et al., 2023; Cornille et al., 2015; Denoirjean et al., 2021; Feurtey et al., 2017; Ruhsam et al., 2019; Schnitzler et al., 2014). While positive fitness effects were observed in one study (Feurtey et al., 2017), changes in land use and management have severely diminished the population size of these wild relatives, making them vulnerable to genetic swamping through introgression (Cornille et al., 2014). This process is facilitated by the obligate outcrossing nature of apple, leading to hybrid percentages of 3% up to 37% in close proximity to apple orchards.

Strawberry: With a relatively recent domestication history of around three centuries, the cultivated strawberry *Fragaria × ananassa* arose as a hybrid of two octaploid parent species (*Fragaria chiloensis* and *Fragaria virginiana*), but has many diploid relatives (Fan & Whitaker, 2024). Diploid genome assemblies and whole genome re-sequencing of hundreds of samples have recently revealed the complex relationships between *Fragaria* diploid relatives and the subgenomes of the octaploid relatives. A combination of *D*-statistics and phylogenetic networks created with PhyloNet demonstrated extensive reticulation among diploid strawberries, including the woodland strawberry *Fragaria vesca*, with the extent of introgression ranging from 0.2-16.4%, averaging 4.1% of the genome across species (Feng et al., 2023). In addition, repeated gene flow has occurred between the octaploid progenitor species prior to domestication (although its extent wasn't quantified) and signals of interploid introgression likely predate polyploidisation (Fan & Whitaker, 2024). While the domesticated *Fragaria × ananassa* wasn't included in these tests, its hybrids origin testifies to the fact these octaploids are interfertile, and breeding efforts have increased the proportion of the *F. virginiana* in the genome (Fan & Whitaker, 2024).

Solanaceae (Solanales)

Eggplant: Page et al. (2019) studied patterns of admixture between the cultivated eggplants *Solanum melongena* and *S. ovigerum* and their wild progenitor *S. insanum* in Asia*.* Analysing close to 5000 SNPs generated by GBS for 73 individuals, they discovered that a group of feral eggplants from India, classified as *S. insanum* and comprising 23% of total wild accessions, displayed high levels of cultivated ancestry in a STRUCTURE analysis, suggesting and admixed origin. This was subsequently backed up with DIYABC, showing 31-40% cultivated ancestry in *S. insanum*, and a TreeMix analysis, confirming a migration edge from cultivated eggplant into one of the feral accessions. Barchi et al. (2023) used 120 thousand SNPs to study global eggplant diversity of more than 3000 accessions, and found multiple migration edges with a TreeMix analysis among European and Asian eggplants, but also from Asian eggplant into *S. insanum*, thereby confirming the findings by Page et al. (2019). Page et al. (2019) also performed a TreeMix analysis and found a similar number of migrations, although neither of the studies reported the migration weight.

Miscellaneous

Coffee (Rubiaceae, Gentianales): Verleysen et al. (2024) used more than eight thousand SNPs generated by GBS to study patterns of admixture among 471 accessions of cultivated and wild *Coffea canephora* in Congo. Using fastSTRUCTURE to compare global ancestry of coffee grown in home gardens with the INERA coffee collection and wild coffee from the Yangambi rainforest, they found that twelve out of 261 wild rainforest individuals (5%) had >10% cultivated ancestry, most of them found in close proximity to home gardens. The authors used the 'HIest' R package to classify these individuals and found the majority of these to be cultivated genotypes, with only two F2 hybrids and one hybrid-cultivated backcross.

Macadamia (Proteaceae, Proteales): O'Connor et al. (2015) used 11 microsatellites to study population structure and parentage of 284 cultivated and wild macadamia individuals in New South Wales (Australia). A STRUCTURE analysis showed the likely hybrid origins of an unspecified number of wild macadamia (*Macadamia tetraphylla*) individuals, with ~25% or ~50% ancestry from cultivated macadamia (*M. integrifolia*), in addition to individuals with 100% cultivated ancestry. Parentage analysis confirmed that 30 plants (10.6% of all wild plants sampled) had a *M. integrifolia* parent; these hybrids were found in 9 out of 10 populations. The results suggest that crop-to-wild gene flow is mediated by both pollen and seed dispersal, the latter of which is possible over long distances by water and animal vectors.

Yam (Dioscorea, Dioscoreales): Scarcelli et al. (2017) used a combination of nuclear markers (12 microsatellites) and chloroplast markers (304 SNPs) to demonstrate that nearly half of 65 wild yams collected in Benin might be of hybrid origin, with genetic contributions from cultivated yam (*D. rotundata*). A STRUCTURE analysis showed that 19% of wild individuals (both *D. praehensilis* and *D. abyssinica*) have >20% of cultivated ancestry. Chloroplast haplotypes subsequently showed that an additional 24% of wild yams, while displaying majority wild ancestry in the nuclear genome, carried the domesticated chloroplast haplotype, consistent with more ancient hybridisation and subsequent purging of the nuclear cultivated ancestry over time. The high frequency of cultivated chloroplast haplotypes in wild yams (46%) indicates that crop-to-wild gene flow is mediated by pollination of cultivated female yams by wild male yams. These could end up in the wild compartment through subsequent seed dispersal, or the common practice of abandonment of agricultural fields associated with slash-and-burn agriculture.

B.2 Studies included in review

Table B.2.1. Literature surveyed and included in this study. For each paper we extracted the names and sample sizes of crop and wild species studied, the location of study, the type of marker used for genotyping and the number of markers for which data were generated, and the analytical method used to study introgression. Analytical methods are separated by the way they quantify introgression: population extent (the percentage of admixed/hybrid individuals in a population, based on a certain threshold of cultivated ancestry), genomic extent (relative to the genome size or in megabase pairs); the number of introgression events (the relative or absolute number of positive tests), population graphs or networks (either expressed as the number of gene flow events/migration edges inferred, as the relative weight (w) or proportion of admixture per gene flow event, or both), or migration rate (expressed as the number of migrants per generation (m), or the mutation-scaled migration rate (M)). Studies on crop-to-wild introgression from cultivated species that are the subject of R&D with NGTs (Table S2) are highlighted in blue.

Article	Study system					Study design			Analysis					
Author (Year)	**Vernacular name**	**Cultivated species**	**Relative species**	**Status**	**Geographic location**	**Sample size**	**Number of markers**	**Type of marker**	**Population extent: global ancestry (% of population)**	**Population extent: hybrid classification (% of population)**	**Genomic extent (% of genome/Mb)**	**Introgression events (positive tests)**	**Graphs and networks (edges, rate)**	**Migration rate (migrants per generation)**
Adugna et al. (2017)	Sorghum	*Sorghum bicolor* subsp. *bicolor*	*Sorghum bicolor*	Wild	Ethiopia	540	12	SSR	STRUCTURE: ?					Fst: M = 2.13
Arrigo et al. (2011)	Wheat	*Triticum aestivum*	*Aegilops geniculata; Ae. Neglecta; Ae. triuncialis*	Wild	Spain, France and Italy	427	191	AFLP	STRUCTURE: 3%; 15%; 26% (>0.1 ancestry)	Fuzzy c-means clustering: consistent with STRUCTURE				
Barchi et al. (2023)	Eggplant	*Solanum melongena*	*Solanum insanum*	Feral	Asia	3499	119695	SNPs (TC)					TreeMix: ?	
Baute et al. (2016)	Sunflower	*Helianthus annuus*	*Helianthus bolanderi-exilis; H. argophyllus*	Wild	North America	292	4645	SNPs (GBS)			fd: 7.1-9.7%			
Bina et al. (2022)	Apple	*Malus domestica*	*Malus orientalis*	Wild	Caucasus	426	26	SSR	STRUCTURE: 11% (>0.1 ancestry)					
Chen et al. (2023)	Apple	*Malus domestica*	*Malus sylvestris*	Wild	Europe	624	13	SSR	STRUCTURE: 12% (>0.1 ancestry)					
Choi et al. (2017)	Rice	*Oryza sativa* subsp. *japonica*	*Oryza sativa* subsp. *indica/O. nivara*	Wild	Asia	5	19000	Genes (WGS)				D + Dfoil: ?	G-PhoCS: 3 events; 11-17%	
Cornille et al. (2013)	Apple	*Malus domestica*	*Malus orientalis; M. sieversii; M. sylvestris*	Wild	Caucasus; Central Asia; Europe	1221	26	SSR	STRUCTURE: 3.2%; 15%; 37% (>0.1 ancestry)					
Cornille et al. (2015)	Apple	*Malus domestica*	*Malus sylvestris*	Wild	Europe	2228	26	SSR	STRUCTURE: 24% (>0.1 ancestry)					
Delplancke et al. (2012)	Almond	*Prunus dulcis*	*Prunus orientalis*	Wild	Lebanon, Turkey, and Syria	562	12	SSR	STRUCTURE: ?					MIGRATE-N: M = 28
Denoirjean et al. (2021)	Apple	*Malus domestica*	*Malus sylvestris*	Wild	Europe	55	13	SSR	STRUCTURE: 10% (>0.1 ancestry)					
Fan et al. (2024)	Strawberry	*Fragaria × ananassa*	various (octaploid)	Wild	n/a	289	1866574	SNPs (WGS)			HyDe: ?	D (Dsuite): ?; f-branch: ?		
Feng et al. (2023)	Wild strawberry	*Fragaria vesca*	various (diploid)	Wild	n/a	69	14175029	SNPs (WGS)			fhom: 0.2-16.4%	Dfoil: ?; D (Dsuite) + fhom: 79% of triplets		

Article	Study system					Study design			Analysis					
Author (Year)	Vernacular name	Cultivated species	Relative species	Status	Geographic location	Sample size	Number of markers	Type of marker	Population extent: global ancestry (% of population)	Population extent: hybrid classification (% of population)	Genomic extent (% of genome/Mb)	Introgression events (positive tests)	Graphs and networks (edges, rate)	Migration rate (migrants per generation)
Feurtey et al. (2017)	Apple	*Malus domestica*	*Malus sylvestris*	Wild	France	186	33	SSR		HI (introgress): 23%				
Feurtey et al. (2020)	Apple	*Malus domestica*	*Malus sieversii*	Wild	Kazakhstan	240	28	SSR	STRUCTURE: 11% (>0.1 ancestry)					
Guo et al. (2022)	Barley	*Hordeum vulgare*	*Hordeum vulgare* subsp. *spontaneum*	Wild	Central and Eastern Asia	212	?	SNPs (WGS)	ADMIXTURE: almost all (>0.2 ancestry)		f4 (ADMIXTOOLS: qpDstat): ?			
Gutierrez et al. (2010)	Sunflower	*Helianthus annuus*	*Helianthus petiolaris*	Weedy	Argentina	26	51	RAPD		Diagnostic alleles: 2%				
Ha et al. (2021)	Apple	*Malus domestica*	*Malus sieversii*	Wild	Kazakhstan	120	17	SSR						MIGRATE-N: m 0.00037-0.0015; BayesAss: m = 0.03-0.06
Helmstetter et al. (2020)	Hazelnut	*Corylus avellana*	*Corylus avellana*	Wild	Turkey	200	60000	SNPs (RAD)				D (PopStat): 1 migration event	TreeMix: ?	
Heredia-Pech et al. (2022)	Lima bean	*Phaseolus lunatus*	*Phaseolus lunatus*	Wild	Yucatan peninsula (Mexico)	83	15168	SNPs (GBS)	STRUCTURE: 36% (>0.1 ancestry)		NGSEP: 94 blocks; f4 (Dsuite): 15-24%; Df (Dsuite): 12-18%			MIGRATE-N: M = 114-157
Hernández et al. (2023)	Carrot	*Daucus carota*	*Daucus carota*	Wild	United States	594	34871	SNPs (GBS)	ADMIXTURE: 24% (>0.1 ancestry)		f4 (Dsuite): 15.2-24.3%	D (Dsuite): 26% of pairs		
Hufford et al. (2013)	Maize	*Zea mays* subsp. *mays*	*Zea mays* subsp, *mexicana*	Wild	Mexico	189	39029	SNPs (array)	STRUCTURE: ?		HAPMIX: 9.2-11.4%	f3: 8/9 populations		
Jin et al. (2024)	Barley	*Hordeum vulgare*	various	Wild	Global	67	103	Genes (TC)				D + Dfoil; QuIBL: 5% of triplets		
Le Corre et al. (2020)	Maize	*Zea mays subsp. mays*	*Zea mays* subsp. *mexicana*	Weedy	France	319	24544	SNPs (array)			ELAI: ?; f4: 12.2-42.2%		TreeMix: w = 0.14-0.39	
Mondon et al. (2018)	Sunflower	*Helianthus annuus*	*Helianthus annuus; H. petiolaris*	Wild	Argentina	182	3526	SNPs (GBS)				D (ANGSD: abbababa2): ?		
Muller et al. (2011)	Sunflower	*Helianthus annuus*	*Helianthus annuus*	Weedy	France and Spain	332	16	SSR + mtDNA	STRUCTURE: ?	Diagnostic alleles (mtDNA): 100%				
Mutegi et al. (2012)	Sorghum	*Sorghum bicolor* subsp. *bicolor*	*Sorghum bicolor; S. bicolor* nothosubsp. *drummondi*	Wild	Kenya	483	10	SSR			STRUCTURE: up to 28%			BayesAss: m = 0.08; 29% of population
O'Connor et al. (2015)	Macadamia	*Macadamia integrifolia*	*Macadamia tetraphylla*	Wild	New South Wales (Australia)	284	11	SSR	STRUCTURE: ?	Parentage analysis (CERVUS): 10.6%				
Olodo et al. (2020)	Pearl millet	*Pennisetum glaucum*	*Pennisetum glaucum*	Wild	Senegal	1222	12	SSR			STRUCTURE: 3%			
Olofsson et al. (2024)	Spearmint	*Mentha spicata*	*Mentha longifolia; M. suaveolens*	Wild	Sweden and other Nordic countries	93	1688670	SNPs (WGS)			NGSadmix: ?	D: 58% of trios		
Omasheva et al. (2017)	Apple	*Malus domestica*	*Malus sieversii*	Wild	Kazakhstan	361	16	SSR	STRUCTURE: 8-95% (>0.1 ancestry)					
Page et al. (2019)	Eggplant	*Solanum melongena*	*Solanum insanum*	Weedy	Asia	73	4880	SNPs (GBS)	STRUCTURE: 23% (no threshold given)				TreeMix: 1-3 events; DIYABC: 31-40%	
Pajkovic et al. (2014)	Wheat	*Triticum aestivum*	*Aegilops triuncialis*	Wild	Southern Spain	491	491	AFLP	InStruct: 10% (>0.1 ancestry)	Fuzzy c-means clustering: consistent with InStruct				ABC: m = 2.7 × 10−4 per resident

Article	Study system					Study design			Analysis					
Author (Year)	Vernacular name	Cultivated species	Relative species	Status	Geographic location	Sample size	Number of markers	Type of marker	Population extent: global ancestry (% of population)	Population extent: hybrid classification (% of population)	Genomic extent (% of genome/Mb)	Introgression events (positive tests)	Graphs and networks (edges, rate)	Migration rate (migrants per generation)
Parisod et al. (2013)	Wheat	*Triticum aestivum*	*Aegilops triuncialis*	Wild	Spain; California (United States)	376	12	SSR		Diagnostic alleles: 17.7% of individuals; 66.6% of populations				
Presotto et al. (2024)	Rice	*Oryza sativa* subsp. *japonica; indica; aus*	*Oryza sativa* subsp. *aus; aus-indica*	Weedy	Argentina	88	15159	SNPs (GBS)				D (Dsuite): ?		
Qiu et al. (2020)	Rice	*Oryza sativa*	*Oryza sativa*	Weedy	Global	1003	16200000	SNPs (WGS)	fastStructure: 54% in Latin America (no threshold given)			D: ?	TreeMix: ?	
Rabanus-Wallace et al. (2021)	Rye	*Secale cereale*	*Secale vavilovii; S. strictum; S. sylvestre*	Wild	n/a	955	72465	SNPs (GBS)				D: 10 pairs		
Rojas-Barrera et al. (2019)	Maize	*Zea mays* subsp. *mays*	*Zea mays* subsp. *mexicana*	Wild	Mexico	385	316294	SNPs (GBS)				D: 12 pairs		
Roumet et al. (2013)	Sunflower	*Helianthus annuus*	*Helianthus annuus*	Weedy	France	766	16	SSR		Parentage analysis: 6.7-38%				
Ruhsam et al. (2019)	Apple	*Malus domestica*	*Malus sylvestris*	Wild	United Kingdom	343	14	SSR	STRUCTURE: 16-27% (>0.1 ancestry)	NewHybrids: consistent with STRUCTURE				
Saban et al. (2023)	Cabbage	*Brassica oleracea*	*Brassica cretica*	Wild	n/a	79	5800000	SNPs (WGS)			fd (Dsuite): 9.46-14.28%		PhyloNet: 3 events	
Scarcelli et al. (2017)	Yam	*Dioscorea rotundata*	*Dioscorea abyssinica; D. praehensilis*	Wild	West Africa	100	12	SSR + cpDNA	STRUCTURE: 19% (>0.2 ancestry)	Diagnostic alleles (cpDNA): 43%				
Schnitzler et al. (2014)	Apple	*Malus domestica*	*Malus sylvestris*	Wild	Rhine Valley	295	15	SSR	STRUCTURE: 4% (>0.3 ancestry)					
Schreiber et al. (2019)	Rye	*Secale cereale*	*Secale cereale* subsp. *vavilovii*	Wild	Europe and Western Asia	86	55744	SNPs (GBS)				D (ADMIXTOOLS): ?		
Shivrain et al. (2010)	Rice	*Oryza sativa* subsp. *japonica*	*Oryza sativa* subsp. *indica*	Weedy	Arkansas (United States)	173	27	SSR	STRUCTURE: 25% (>0.2 ancestry)					
Spear et al. (2023)	Sunflower	*Helianthus annuus*	*Helianthus annuus*	Wild	Minneapolis (United States)	39	122000	SNPs (RNA)			STRUCTURE: 7.1-13.6%			
Uwimana et al. (2012)	Lettuce	*Lactuca sativa*	*Lactuca serriola*	Weedy	Europe	7738	10	SSR	STRUCTURE: 7% (>0.1 ancestry)	NewHybrids: consistent with STRUCTURE				
Verleysen et al. (2024)	Coffee	*Coffea canephora*	*Coffea canephora*	Wild	Congo	471	8131	SNPs (GBS)	STRUCTURE: 5% (>0.1 ancestry)	HI (HIest): 23 (of 70 admixed) individuals				
Wang et al. (2019)	Soybean	*Glycine max*	*Glycine soja*	Wild	Global	302	?	SNPs (WGS)			rIBD: 0.059-41% (average 1.9%)	D: lower for regions with high rIBD		
Wang et al. (2023)	Rapeseed	*Brassica napus*	*Brassica oleracea*	Cultivated and wild	Global	413	4287347	SNPs (WGS)			f4: 6-22%	D: 15 pairs	TreeMix: many events	
Wedger et al. (2022)	Rice	*Oryza sativa* subsp. *japonica*	*Oryza sativa*	Weedy	Arkansas (United States)	97	19340000	SNPs (WGS)	ADMIXTURE: 73% (>0.15 ancestry)		Loter: 26-31%			
Wu et al. (2023)	Rice	*Oryza sativa* subsp. *japonica*	*Oryza sativa* subsp. *indica/O. nivara*	Wild	n/a	74	175528	Syntelogs (WGS)			fd: 21.2 Mbp			
Zunino et al. (2024)	Olive	*Olea europaea*	*Olea europaea*	Wild	Western Mediterranean Basin	520	142060	SNPs (TC)					TreeMix: w = 0.425-0.485	