

Published by:

Springer Nature Switzerland AG 2018

Citation:

Miller RE and Khoury CK (2018) “The Gene Pool Concept Applied to Crop Wild Relatives: An Evolutionary Perspective”. In: Greene SL, Williams KA, Khoury CK, Kantar MB, and Marek LF, eds., *North American Crop Wild Relatives, Volume 1: Conservation Strategies*. Springer, doi: 10.1007/978-3-319-95101-0_6.

https://doi.org/10.1007/978-3-319-95101-0_6

https://link.springer.com/chapter/10.1007%2F978-3-319-95101-0_6

Chapter 6

The Gene Pool Concept Applied to Crop Wild Relatives: An Evolutionary Perspective

Richard E. Miller* and Colin K. Khoury

Richard E. Miller, 407 East Charles Street, Hammond, Louisiana 70401, rickeltonmiller@gmail.com

*corresponding author

Colin K. Khoury, USDA, Agricultural Research Service, Center for Agricultural Resources Research, National Laboratory for Genetic Resource Preservation, Fort Collins, CO USA and International Center for Tropical Agriculture (CIAT), Cali, Colombia, colin.khoury@ars.usda.gov c.khoury@cgiar.org

Abstract Crop wild relatives (CWR) can provide important resources for the genetic improvement of cultivated species. Because crops are often related to many wild species and because exploration of CWR for useful traits can take many years and substantial resources, the categorization of CWR based on a comprehensive assessment of their potential for use is an important knowledge foundation for breeding programs. The initial approach for categorizing CWR was based on crossing studies to empirically establish which species were interfertile with the crop. The foundational concept of distinct gene pools published almost 50 years ago was developed from these observations. However, the task of experimentally assessing all potential CWR proved too vast; therefore, proxies based on phylogenetic and other advanced scientific information have been explored. A current major approach to categorize CWR aims to comprehensively synthesize experimental data, taxonomic information, and phylogenetic studies. This approach very often ends up relying not only on the synthesis of data but also intuition and expert opinion, and is therefore difficult to apply widely in a reproducible manner. Here, we explore the potential for a stronger standardization of the categorization method, with focus on evolutionary relationships among species combined with information on patterns of interfertility between species. Evolutionary relationships can be revealed with increasing resolution via next-generation sequencing, through the application of the multispecies coalescent model and using focused analyses on species discovery and delimitation that bridge population genetics and phylogenetics fields. Evolutionary studies of reproductive isolation can inform the understanding

of patterns of interfertility in plants. For CWR, prezygotic postpollination reproductive barriers and intrinsic postzygotic barriers are the most important factors and determine the probability of producing viable and fertile offspring. To further the assessment of CWR for use in plant breeding, we present observed and predicted gene pool indices. The observed index quantifies patterns of interfertility based on fertilization success, seed production, offspring viability, and hybrid fertility. The predicted gene pool index requires further development of the understanding of quantitative and qualitative relationships between reproductive barriers, measures of genetic relatedness, and other relevant characteristics for crops and their wild relatives.

Keywords Multispecies coalescent model · Postzygotic reproductive barriers · Reproductive isolation · Solanaceae · Species delimitation · Targeted gene capture · Unified species concept.

6.1 Introduction

Everyone interested in biological diversity has a sense of what a species is. We recognize collections of individual organisms that look similar and give them common names, distinguishing these individuals from other dissimilar groups. For example, the bald eagle (*Haliaeetus leucocephalus* L.) is a North American sea eagle (sea eagles: Haliaeetinae) with a white head and tail and a brown body; it is a very distinctive large raptor within its range. It is easy to recognize as a different species from the larger golden eagle (*Aquila chrysaetos* L.) (booted eagles: Aquilinae) (Lerner and Mindell 2005).

A familiar definition of a species is the biological species concept, developed by the ornithologist and evolutionary biologist, Ernst Mayr (1942). The concept states “species are groups of actually or potentially interbreeding natural populations, which are reproductively isolated from other such groups.” The focus of this influential definition (discussed by de Queiroz 2005) is whether populations interbreed or not, or in a more abstract sense, whether there is gene flow between diverging lineages (Petit and Excoffier 2009). This definition can certainly be applied to bald eagles and golden eagles. Even though these birds of prey are sympatric in some areas, they do not interbreed. It is interesting that the biological species concept makes no mention of whether species are morphologically distinct, even though naturalists’ identification of species are most often based on visible characteristics (Coyne and Orr 1998).

Among flowering plants, it is common to observe groups of closely related species that are difficult to distinguish. These confusing plants also may hybridize and produce intermediate forms in the wild (Stebbins 1950; Grant 1981). Many species closely related to crops appear to belong to hybridizing complexes (Small 1984; Ellstrand et al. 1999; Rieseberg et al. 2007). This ability to produce viable offspring from interspecific hybrids can be advantageous to plant breeders. Owing to the interfertility between crops and their crop wild relatives (CWR), the genetic variation among these resources can be harnessed to introduce traits into crops to meet challenges, including pathogens, stressful environments, and changes in management practices (Hajjar and Hodgkin 2007).

Even species that are phenotypically distinct and reproductively isolated from crops may be considered CWR. Research has demonstrated that for many plant species, barriers acting early in the speciation process contribute most to reproductive isolation among distinct species (Martin and Willis 2007; Lowry et al. 2008). Artificial crosses between crops and their wild relatives can

circumvent these early-acting barriers, leading to the successful production of viable hybrid offspring.

Crops are domesticated plant species (Brozynska et al. 2016) or sometimes complex combinations of more than one species, especially if they are of polyploid origin (e.g., bread wheat (*Triticum aestivum* L.) [International Wheat Genome Sequencing Consortium 2014]). An important goal in plant breeding is to identify the crop gene pool (Vincent et al. 2013), which includes the cultivated forms of the crop species, the wild species of the crop, and the crop's progenitors (Brozynska et al. 2016). Also included within the crop gene pool are closely related species, even if they did not serve as direct progenitors.

For some crops, there may be dozens or even hundreds of related congeneric species, for example, in the genus *Solanum* L. (Vorontsova et al. 2013; Hardigan et al. 2015). It is therefore critical to be able to discern which species are likely to be most successful in plant breeding efforts. To accomplish this, both a foundational knowledge of the evolutionary relationships among crops and their wild relatives and a comprehensive understanding of the relationship between patterns of interfertility and genetic relatedness and reproductive characteristics of CWR are needed.

6.2 A History of the Gene Pool Concept for CWR

Crop wild relatives are not merely the relatives of crops. All angiosperms share a common ancestor; therefore, all angiosperms are related. The practical implication of being defined as a CWR is that a plant species has some potential to be useful in a breeding program (Harlan and de Wet 1971). There is no taxonomic or evolutionary entity that corresponds to the notion of CWR rather it is a concept specific to crop science and plant breeding.

The two main uses of the gene pool concept are to delineate the taxa that are CWR and then to identify levels of interfertility between these taxa and the associated crop. The early and very influential system for recognizing gene pools was developed by Harlan and de Wet (1971). They proposed three tiers of gene pools. The first tier corresponded to the biological species including the crop and, therefore by definition, individuals that exhibited no barriers to reproduction. The second tier included other taxa that could cross with the crop but with difficulty. Harlan and de Wet (1971) alluded to what would now be considered linkage drag, the retention of undesirable traits in the crop-wild relative hybrid (Prohens et al. 2017), making the use of these materials feasible but problematic. Finally, they recognized a tertiary group of taxa that were at the limits of interfertility, which could produce a few anomalous seeds or could be utilized only via radical techniques such as embryo rescue (Jansky 2006). Harlan and de Wet (1971) struggled to reconcile the biological diversity included within cultivated plants and their relatives with classic taxonomic categories, especially at the subspecific level, realizing that equivalent gene pool categories were unlikely.

The data required to make the assessment of placing CWR into Harlan and de Wet's gene pool tiers come from crossing studies to establish levels of interfertility. These are very challenging experiments to carry out and require substantial effort. A recent study of reproductive barriers among species of *Jaltomata* Schltdl. (Solanaceae) (Kostyun and Moyle 2017) provides an excellent example of the magnitude of what is needed to determine levels of interfertility among species. The authors investigated 10 species and performed crosses in a full diallel with typically 6 or 7 individual pairs for each interspecific cross and 8 - 10 crosses per

pair, resulting in 4,481 crosses. The study included additional control crosses (the number of which depended on the breeding system of the species involved—self-compatible versus self-incompatible).

These crossing experiments required a large greenhouse space and a significant investment in time. Furthermore, such controlled cross-pollinations require careful manipulations of the flowers, such as emasculations of the dams, which is tedious (Eserman 2012). Given the tremendous effort needed to determine levels of interfertility among groups of species (Bohs 1991; Smith and Baum 2007; Eserman 2012; Plazas et al. 2016), it is thus not surprising that the comprehensive experiments required to confidently assign species to different tiers of the gene pool are often lacking for CWR, particularly for those related to non-major staple crops.

Proposals have been forwarded to use alternative sources of information to assign CWR to gene pools. One system is to use the existing taxonomies for species to assign CWR to gene pools, with the assumption that taxonomic classification in part reflects levels of genetic relatedness (Maxted et al. 2006). The taxon groups proposed by Maxted et al. (2006) included five taxon groups: the crop and its wild conspecifics, followed by members of the same series or section, subgenus, genus, and finally by different genera in the same tribe. The benefit of using the taxon group for designating CWR to different tiers is that more species can be considered with this scheme without requiring extensive experimental data. Unfortunately, this is a very blunt tool for assigning CWR to categories that are meant to reflect levels of interfertility. For example, generic status should reflect taxon group 4. Considering two solanaceous genera, *Solanum* L. with approximately 1,500 species (Weese and Bohs 2007) and *Iochroma* Benth. with 24 species (Smith and Baum 2007), it seems highly unlikely that all of the species within each genus represent the same level of phylogenetic relatedness to the other members.

One reason genera are not a good proxy for levels of interfertility among species is that taxonomic categories above the species level are arbitrary and therefore often reflect the history of the treatment of the group rather than genetic relatedness. For example, within the tribe that represents morning glories with spiny pollen, the Ipomoeae, the Old World treatment of morning glories by Verdcourt (1963) retained earlier circumscribed genera within the tribe (e.g., *Argyreia* Lour., *Stictocardia* Hallier f.), while in the New World treatment of the morning glories, Austin (1997) dissolved previously recognized genera within the broad and cumbersome genus *Ipomoea* L. (e.g., *Batatas* Choisy, which contains sweetpotato, and *Pharbitis* Choisy). Put another way, levels of interfertility are not uniform for a particular taxonomic level. For example, among 11 close relatives of sweetpotato (members of *Ipomoea* series *Batatas* (Choisy) D.F. Austin), levels of interfertility ranged from very high to extremely low (Diaz et al. 1996), whereas among different genera within the Iochrominae, Smith and Baum (2007) reported high levels of interfertility throughout the tribe. Another important observation is that rates of speciation are not uniform across lineages, with some genera capturing high levels of diversity of recent rapid radiations (Kay et al. 2005). Thus, the taxon group proposal, while making an important effort to overcome the limitation set by reliance on hybridization evaluations, does not provide an accurate indicator of interfertility that can be used confidently.

Wiersema and León (2016) recognized both the pragmatic challenges of the Harlan and De Wet (1971) gene pool concept and the limitations of a system emphasizing taxonomic hierarchies (Maxted et al. 2006). They proceeded to marshal all available information, including taxonomic investigations combined with phylogenetic analyses and genetic and breeding studies, to delineate CWR and place them in categories of genetic relatedness as an updated approximation of the Harlan and de Wet (1971) gene pools. With this system, Wiersema and

León designated the CWR of 208 crops from within 109 genera, based on an exhaustive review of the literature (J. Wiersema, pers. comm.). In the subsequent sections, we outline some of the most important concepts, methods, and sources of data used to inform an updated gene pool concept starting with consideration of the fundamental unit of CWR, the species.

6.3 What Is a Species?

Characterizing the complex relationships commonly observed among CWR requires a more encompassing notion of a species than the biological species concept (Levin 1979; Luckow 1995; de Queiroz 2005). Fortunately, after decades of discussion, a unified species concept has been successfully developed (Hey 2006; De Queiroz 2007). An important foundation of this idea is the evolutionary species concept, wherein species are considered separately evolving metapopulation lineages, which are populations with ancestor-descendant relationships that share the same evolutionary trajectory.

The important challenge then is to take this unified conceptual view of an evolutionary species and develop an operational approach that allow taxa to be recognized and delineated. The unified species concept provided the needed synthesis bringing together various operational species definitions under a single umbrella (De Queiroz 2007). The insight of de Queiroz (2007) was that a suite of properties could be extracted from the many species definitions and integrated, recognizing speciation as an extended process.

Two populations of a single species can be envisioned to diverge over time into two lineages, finally resulting in two clearly separate species. The speciation process is an extended one, where a series of steps cause the daughter lineages to differ to an ever-greater degree (de Queiroz 2007). Between the pairs of lineages, reproductive barriers develop during the speciation process, but not in the same sequence for all species. Diverging lineages may develop different fertilization systems; become adapted to different niches; accumulate random mutations increasing genetic differences; develop diagnosable, fixed differences; become phenetically distinguishable or other changes. Furthermore, a phylogenetic analysis would find the lineages to be reciprocally monophyletic. An examination of gene genealogies for the different lineages would detect some gene histories exhibiting exclusive coalescence of alleles (de Queiroz 2007). At what point along the speciation continuum a researcher defines speciation to have occurred may depend on their area of expertise. From a genetic modeling point of view, perhaps 5% gene flow may be a useful operational criterion (Wiens and Servedio 2000). From a phylogenetic perspective, monophyly might be emphasized (Donoghue 1985). However, a broad approach using all evidence is now considered to be the most informative, allowing the data to tell the speciation story without a preconceived notion of which properties are definitive (de Queiroz 2007; O'Meara 2009; Fujita et al. 2012; Carstens et al. 2013).

Traditional taxonomic research emphasizes morphological characteristics to assign populations to described species. The treatment of *Ipomoea* in Bolivia (Wood et al. 2015) provides an elegant example of a morphology-based taxonomy. Descriptions and dichotomous keys were provided for 102 Bolivian morning glories species, with distribution maps, line drawings, and photographs for select species. Eighteen novel species were introduced, including a new member of the *Batatas* group, the closest relatives of sweetpotato (*Ipomoea batatas* (L.) Lam.).

But, when morphological analyses are combined with genetic data, as well as experimental crosses, a more complicated picture of relationships can emerge. The CWR of sweetpotato are considered members of the *Batatas* group, a hybridizing complex of 16 species distributed from Central United States to northern Argentina (Khoury et al. 2015). In an investigation that represents a snapshot of this complex, 154 Carolina populations representing four species (*Ipomoea cordatotriloba* Dennst., *I. lacunosa* L., *I. leucantha* Jacq., and an undescribed species referred to as “*I. austinii*” [non *I. austinii* Infante-Bet.]), revealed that *I. cordatotriloba* and “*I. austinii*” were morphologically distinct from the other taxa, while *I. lacunosa* and *I. leucantha* were morphological similar, although with white and purple corollas, respectively (Duncan and Rausher 2013a, b). In contrast, *I. cordatotriloba* and *I. lacunosa*, while morphologically distinct, were not genetically differentiated and only exhibited partial reproductive isolation, whereas “*I. austinii*” and *I. leucantha* were genetically distinct. These complicated patterns of evolutionary relationships are consistent with the expectations of the early stages of the expanded speciation process. A more definitive species delineation for the CWR of sweetpotato thus requires further work to integrate traditional and newer tools to better discern the identities of taxa, with a view of their status within the speciation process.

CWR of other crops would be similarly expected to exhibit characteristics of taxa early in speciation. Among the collective populations and species represented by such groups, there may be population structure related to geographic barriers to gene flow, species diagnosed based on subtle morphological features, gene flow that may be ongoing between diverging lineages, and introgressive hybridization that may be occurring. These complications may all occur simultaneously among a group of CWR.

Thankfully, the potential to untangle complex evolutionary relationships and determine species boundaries is ever higher, due to the updated synthesis definition of a species, including the useful conceptual framework for the speciation process. In addition, excellent sources of genetic data and a rapidly developing toolkit of analytical methods are at researchers’ disposal. A brief description of the most important recent technical, theoretical, and analytical advances is provided below.

6.4 Genetic Data for Understanding Evolutionary Relationships Among CWR

Genetic analysis techniques, specifically next-generation sequencing (NGS), can provide a wealth of DNA sequence data useful for examining evolutionary relationships (McCormack et al. 2013; Soltis et al. 2013). These methods have made it feasible, with a concerted effort and great expertise, to sequence entire nuclear genomes (Albert et al. 2013). For example, the complex and massive hexaploid bread wheat genome was recently sequenced (International Wheat Genome Sequencing Consortium 2014). It was particularly challenging because the polyploid AABBDD structure of the genome meant that three pairs of each gene exist. Furthermore, the wheat genome is riddled with repeats from transposable elements, making genome assembly extremely complicated. The near-complete sequence of the wheat genome was developed through a combination of high coverage with short reads from Illumina sequences and long reads from Pac Bio sequences. Approximately 100,000 CPU hours of computer time were

required to stitch together the pieces to assemble 15,343,750,409 base pairs into a single genome (Zimin et al. 2017).

Whole nuclear genome sequences, especially for crops, are being completed at an amazing pace; at least 30 crop species have been sequenced (Morrell et al. 2012; Brozynska et al. 2016). The rice genome is notable because of its simplicity and has been a model for gene identification. In addition, although it is not a crop, the vast information for the model organism *Arabidopsis thaliana* (L.) Heynh. provides the foundation for the understanding of the molecular genetics of plants (The Arabidopsis Genome Initiative 2000).

Given these advances, quickly and inexpensively obtaining nuclear genome sequences of CWR species to determine evolutionary relationships may someday be possible, although this is not presently feasible (Soltis et al. 2013). Luckily, various methods are being developed that provide valuable information based on a reduced genome sample (Mascher et al. 2013; McCormack et al. 2013; Hirsch et al. 2014). These more limited samples can be obtained by genome skimming, restriction site associated DNA sequencing, and targeted enrichment/gene capture. The strengths and weaknesses of these methods have been reviewed elsewhere (Hirsch et al. 2014; Heyduk et al. 2016).

It is valuable to consider the gene capture approach in detail, as this technique may be most informative at population and species levels, making it particularly suitable for untangling the complex relationships among CWR (McCormack et al. 2013; Lemmon and Lemmon 2013; Weitemier et al. 2014; Grover et al. 2015). The foundation of this method for estimating evolutionary relationships is to obtain sequences for sets of orthologous genes for the taxa under consideration. This is a major challenge because many genes are members of large and complex gene families (Morrell et al. 2012).

A set of probes (e.g., RNA baits) is designed to target the orthologous genes and to selectively sequence portions of the genome. Ideally, a reference nuclear genome is available among the taxa sampled or for a closely related species. Transcriptome sequences also can be used to develop the probe set, and for this the resources of the OneKP transcriptome project (<http://www.onekp.com>) are invaluable. At the simplest level, probes can be designed with a collection of expressed sequence tags (ESTs), as was successfully done in a study of the phylogenetics of the Asteraceae (Mandel et al. 2014).

After designing a set of probes, these RNA baits are hybridized to the DNA of the taxa included in the study. The hybridized samples are withdrawn from the pool of probes and DNA and are sequenced using a high-throughput sequencer (Illumina). The reads from the selective sequencing are meant to include only orthologous loci, but postsequencing filtering is needed. For example, the reads can be mapped onto the exon reference sequences to confirm that only orthologs are included in the analyses. The aligned sequences from these efforts can then be included in a wide array of population genetic and phylogenetic analyses.

The phylogenetic investigation of the palm genus *Sabal* Adans. by Heyduk et al. (2015) provides a good example of the elements of a targeted enrichment research project. This study examined relationships among a recalcitrant group of 15 species with interesting biogeography. To apply the gene capture approach, a set of 120-bp RNA baits was developed, benefiting from the availability of nuclear genome sequences, specifically from the date palm, *Phoenix dactylifera* L., and the African oil palm, *Elaeis guineensis* Jacq. (Heyduk et al. 2015). In addition, information from three transcriptomes assemblies (*Cocos nucifera* L., *Nypa fruticans* Wurmb., and *Sabal bermudana* L.H. Bailey) were available from the OneKP project (www.onekp.com). Together, this information was used to identify 837 exon sequences from

176 nuclear genes. The baits were hybridized with DNA from all 15 *Sabal* species, as well as two outgroup taxa. The sequences captured (both exon and intron) ranged from 48,965 to 355,729 base pairs. On average 159 of the 176 targeted nuclear genes were obtained among the taxa. The multiple-gene dataset was then analyzed using three coalescent-based phylogenetic methods (STAR [Liu et al. 2009]; MP-EST [Liu et al. 2010]; *BEAST 1.7.5 [Heled and Drummond 2010]). A well-resolved species tree was recovered that was largely congruent among the different coalescent-based methods. The biogeography of these palms also was resolved, identifying a Central and South American clade, a United States clade, and a clade of mostly Caribbean species. However, recent relationships remained ambiguous even with the large multigene dataset. Population-level sampling will be needed to understand the cause of this poor resolution.

It would be ideal to have a set of probes available to generate multigene datasets of orthologous loci that could be used for all CWR. But the magnitude of this endeavor is daunting, given the number of species of interest. For example, Vincent et al. (2013) and Dempewolf et al. (2014) estimated there may be thousands of potential CWR species. Animal studies have made use of an almost universal set of ultra-conserved elements (UCEs) that can be employed across a wide sample of faunal diversity (Faircloth et al. 2012). An effort is underway to identify a universal set of orthologous conserved loci in flowering plants (Buddenhagen et al. 2016). However, whole-genome duplication events are common across angiosperms, creating complicated genomes and important genetic differences among lineages (Duarte et al. 2010; Albert et al. 2013; Weitemier et al. 2014). A more feasible alternative is to develop a probe set for each plant order (Soltis et al. 2013). Nuclear plant genomes are already available for 30 crop species that represent 13 plant orders and as such could provide a foundation. However, a question that remains is how far-reaching a single set of probes can be. Comer et al. (2016) demonstrated that the probe set Heyduk et al. (2015) designed for the palm genus *Sabal* could be applied more broadly to explore phylogenetic relationships among exemplars of the palm subfamily Arecoideae. However, could RNA baits specifically designed for the CWR of potato (*Solanum tuberosum* L.), for example, be equally useful for resolving fine-scale relationships among the CWR of other members of the Solanaceae, such as tomato (*Solanum lycopersicum* L.), eggplant (*Solanum melongena* L.), chili pepper (*Capsicum annuum* L.), and tobacco (*Nicotiana tabacum* L.)? The answer to this question awaits further research.

6.5 Theoretical Foundation for Understanding Evolutionary Relationships Among CWR

The coalescent model is a stochastic process that can be used to explore relationships among populations, as well as among closely related species, making it excellent for understanding the evolution of CWR. DNA sequence polymorphism data are used to evaluate a wide range of population genetic processes simulated under the coalescent model. A scenario of particular relevance to CWR is when two populations diverge in isolation and where there is no gene flow between them. Population-level divergence of the coalescent process is therefore extended to provide a simple model of speciation or the multispecies coalescent. This model can be used to estimate evolutionary relationships, as well as accommodate other complicated processes, such as migration.

Taking a step back, an important model of genes evolving in populations is the Wright-Fisher model of genetic drift (Wakeley 2009). In the simplest case, the frequencies of two alleles at a single locus change from one generation to the next (for a haploid organism, without overlapping generations). Lineages are followed forward through time, and they branch when an individual randomly produces two or more offspring; lineages end when individuals do not reproduce. The result is random changes in gene frequencies from one generation to the next. In the absence of natural selection and mutation, owing to finite population size and stochasticity, one allele will eventually become fixed, while the other is lost. The dynamics of polymorphic loci in this hypothetical population provide a null model that can be used to examine how other evolutionary processes may modify patterns of polymorphisms.

The coalescent model is a significant extension of the Wright-Fisher model (Degnan and Rosenberg 2009). This stochastic model of gene evolution incorporates the insight of considering the dynamics of gene histories going backwards in time. In coalescent theory, genetic polymorphisms are the result of genealogical and mutational histories (Nordborg 2001; Rosenberg and Nordborg 2002). Lineages coalesce when individuals are produced by the same parent. For a population of individuals, ancestry is traced back through time, with the number of lineages decreasing, until a single individual is reached, representing the most recent common ancestor (MRCA) (Nordborg 2001). This pattern of gene genealogy provides the foundation for the coalescent approach, with neutral allelic variants mapped onto the genealogy, providing a separate description of the pattern of mutation.

The coalescent theory shares the same assumptions of the Wright-Fisher model (discrete generations, constant effective population size within populations, no population structure, no selection) (Rosenberg and Nordborg 2002; Degnan and Rosenberg 2009). This approach provides a springboard for examining how other processes affect the genealogy, such as population subdivision, species migration, hybridization, horizontal gene transfer between species, recombination, changes in population size, and geographic structure (Nordborg 2001; Degnan and Rosenberg 2009). More specifically, simulations model how the genealogy may vary because of different processes, without concern with the random process of mutation. However, the pattern of mutation among individuals carries the empirical information that allows the underlying unobserved genealogy to be compared to the simulated scenarios (Nordborg 2001).

For the study of crops and their wild relatives, the primary interest is an expanded version of the coalescent approach -- the multispecies coalescent (Liu et al. 2015). In this model, the pattern of mutations and gene genealogies is maintained, and the species tree is added. For example, consider two descendant populations that split and no longer exchange genes with each other. In essence, there are two independent coalescent processes in the diverging lineages. In this scenario, divergence time can be estimated, realizing that all coalescent events must occur in the ancestor. This model of divergence without gene flow can be considered a null model of the splitting of species, providing the bridge between coalescent theory and phylogenetics (Wakeley 2009, 2013).

A fundamental realization with the coalescent approach is that the gene genealogies, or gene trees, are not the same as species trees (Edwards 2009). It is now well established that equating a gene tree with a species tree can lead to misleading inferences (Degnan and Rosenberg 2009). The focus of the coalescent approach is to understand the dynamics of multiple gene trees, nested within the context of a species tree. The signal from these gene trees provides the basis for an exploration of the full range of dynamics among populations and

species while paying attention to the processes that give rise to gene tree discordance (incomplete lineage sorting, horizontal gene transfer, hybridization, natural selection, gene duplication) (Degnan and Rosenberg 2009).

6.6 Species Delimitation of CWR

Two major goals of systematics are to discover and describe species and then determine the phylogenetic relationships among taxa (Wiens 2007; O'Meara 2009). Species discovery using genetic data requires multiple population samples for each taxon. Carstens et al. (2013) provide an excellent review of available methods for species discovery (e.g., Structurama, Gaussian clustering, general mixed Yule coalescent model, O'Meara's heuristic method). An important point of their discussion is that each approach has assumptions that may be violated in some way by the real situation of the study system. Therefore, a suite of analyses should be explored, with a final assessment of species boundaries made from a careful synthesis of the results. It is useful to remember that not all diverging populations are expected to clearly exhibit the properties that demarcate separate species.

For the purposes of this discussion, we will consider STRUCTURE, a Bayesian model-based algorithm that is one of the more popular methods of species discovery (Pritchard et al. 2000; Hubisz et al. 2009). A STRUCTURE analysis for a diploid species employs polymorphic alleles at numerous unlinked loci. Samples of individuals are assigned to clusters based on allele frequencies under Hardy-Weinberg disequilibrium (Fujita et al. 2012; Carstens et al. 2013). The number of clusters (k) supported by the data is evaluated in a series of analyses with different numbers of k (Evanno et al. 2005). These clusters are an estimate of the number of lineages supported by the data, and then individuals are assigned to the lineages with varying levels of confidence.

STRUCTURE analysis can employ a model with or without admixture (Falush et al. 2003). Including admixture allows individuals to be assigned to multiple lineages and therefore provides an estimate of the level of gene flow occurring between them. If the lineages are interpreted to be separate species, this analysis provides an examination of the strength of species boundaries. A strong example of application of STRUCTURE for species delineation is the investigation of relationships among brinjal (*Solanum melongena*), scarlet (*S. aethiopicum* L.), and gboma (*S. macrocarpon* L.) eggplants and 14 species of their wild relatives (Acquadro et al. 2017). Acquadro et al. (2017) used next-generation sequencing to identify 75,399 polymorphic sites among 76 individuals. In their initial analysis of this large dataset, Acquadro et al. (2017) used fastSTRUCTURE (Raj et al. 2014) to identify four major subgroups, including one that included brinjal eggplant, its wild progenitor, and its close relatives (*S. melongena* with *S. insanum* L., *S. incanum* L., *S. linnaeanum* Hepper & P.-M.L. Jaeger, and *S. lichtensteinii* Willd.). However, individuals identified as *S. campylacanthum* Hochst. ex A. Rich., *S. lidii* Sinding, *S. tomentosum* L., *S. vespertilio* Aiton, and *S. violaceum* Ortega showed evidence of admixture, exhibiting membership in as many as three distinct lineages. Following the initial analysis for population genetic structure, closer examination of the subgroups was carried out with additional explorations, using fastSTRUCTURE to identify genetically based clusters within these subgroups.

Following species discovery, the next step is species validation. Again, Carstens et al. (2013) advocate applying multiple methods to the data (BP&P, spedeSTEM). Bayesian

Phylogenetics and Phylogeography (BP&P) provides an example of this process (Yang and Rannala 2010; Rannala and Yang 2017). This is a Bayesian inference procedure for multilocus sequence data, based on the multispecies coalescent. The analysis typically begins with a guide tree, and then various scenarios varying in species number and relationships (topologies) are explored using a Markov chain Monte Carlo search. The maximum number of species is specified at the start, and then posterior probabilities for the nodes of the species tree are obtained, using the sequence data. In other words, BP&P evaluates the strength of various possibilities of species relationships and number of species. The outcome of these analyses is identified lineages. Comparing BP&P to other validation approaches suggests this method tends to oversplit genetic relationships (Carstens et al. 2013). Therefore, it is important to integrate the genetic results with information from biogeography, ecology, morphology, and, when possible, the degree of reproductive isolation to make a synthetic assessment of the number of species and their relationships. With careful examination of the number of species included in the population samples, and their validated relationships, one can proceed with greater confidence to examine phylogenetic relationships among taxa. Various phylogenetic methods have been developed based on the multispecies coalescent model that can complete this final step (ASTRAL, Mirarab and Warnow 2015; *BEAST 1.7.5; Heled and Drummond, 2010; MP-EST, Liu et al. 2010; STAR, Liu et al. 2009).

An example of this integrative approach of combining a BP&P analysis with morphological and ecogeographic data is the study of group of rare North American orchids, the *Corallorhiza striata* Lindl. complex (Barrett and Freudenstein 2011). In this taxonomically challenging group, relationships were examined among two species, *C. bentleyi* Freudenst. and the widespread *C. striata* Lindl. The results based on genetic data, morphology, and geography were not congruent. BP&P resolved four lineages: *C. bentleyi* + *C. striata* var. *involuta* (Greenm.) Freudenst. and three distinct populations of *C. striata* from different regions in North America. In contrast, morphological and geographic evidence supported the separation of *C. bentleyi* and *C. striata* var. *involuta*. The authors settled on recognizing three species, *C. bentleyi*, *C. involuta*, and a widespread *C. striata* s.s. The orchid study not only demonstrates the application of BP&P in combination with morphological and biogeographic analyses but also provides an example of the extended speciation process where different properties were found to delineate various combinations of taxa.

These analyses offer new options to resolve long-standing challenges with regard to CWR gene pools. Returning to the CWR of sweetpotato, the *Batatas* group represents a hybridizing complex, which has challenged taxonomic efforts (Austin 1978; Wood et al. 2015). The characteristics and status of the CWR of sweetpotato may be common for other groups of CWR (e.g., the following are crops and their wild relatives: alfalfa [*Medicago sativa* L.], barley [*Hordeum vulgare* L.], beans [*Phaseolus vulgaris* L.], cassava [*Manihot esculenta* Crantz], carrot [*Daucus carota* L.], cotton [*Gossypium hirsutum* L. and *G. barbadense* L.], eggplants [*Solanum melongena*, *S. aethiopicum*, *S. macrocarpon*], maize [*Zea mays* L. subsp. *mays*], potato [*Solanum tuberosum*], rice [*Oryza sativa* L. and *O. glaberrima* Steud.], sorghum [*Sorghum bicolor* (L.) Moench], soybean [*Glycine max* (L.) Merr.], sugarcane [*Saccharum officinarum* L.], sunflower [*Helianthus annuus* L.], and wheat [*Triticum aestivum*]) where species are diagnosed on disputable characteristics, complex patterns of reproduction have been detected, incomplete barriers to reproduction are present, and cryptic species have been discovered (Small 1984; Ellstrand et al. 1999; Rieseberg et al. 2007; Grover et al. 2015; Hardigan et al. 2015; Bredeson et al. 2016; Acquadro et al. 2017).

6.7 What Determines Patterns of Interfertility Among Plants?

An evolutionary perspective can inform the understanding of patterns of interfertility among CWR from an investigation of reproductive isolation in plants (Lowry et al. 2008; Baack et al. 2015). Determining patterns of reproductive isolation is at the foundation of understanding speciation. As discussed above, the framework of the extended speciation process portrays a view where along the speciation continuum different barriers develop until species are completely reproductively isolated from one another and thereby exhibit an increasing number of the properties that characterize separate species.

To better understand reproductive isolation, different stages of reproductive barriers have been recognized. One of the main distinctions is made between barriers that operate before or after zygote formation (Coyne and Orr 1998). Prezygotic isolating mechanisms involve such factors as differences in ecogeography or elaborate floral structures that result in adaptations to alternative pollinators that act to eliminate gene flow between species pairs before mating takes place (Stebbins 1950). In contrast, postzygotic isolating mechanisms involve factors such as hybrid inviability, where the zygote does not develop into a normal embryo (Stebbins 1950). For plants, a further distinction is made between prezygotic barriers that act prior to pollination (e.g., differences in flowering time) and prezygotic postpollination barriers (e.g., interactions between pollen and stigma) (Tiffin et al. 2001).

An important model for understanding the development of reproductive barriers is the Bateson-Muller-Dobzhansky model or Dobzhansky-Muller incompatibilities (DMIs) model, where allelic incompatibilities between diverging species increase, leading to hybrid inviability and sterility (Coyne and Orr 1998; Tiffin et al. 2001; Turelli and Moyle 2007). The model is based on mutations at two or more loci developing in two diverging populations. If the two populations interbreed after mutations have accumulated, some of the genetic combinations will be incompatible. One prediction from the DMI model is that the level of reproductive isolation between species pairs is positively associated with genetic distance, which is thought to reflect time since divergence. In other words, with greater divergence, reduced hybrid fitness is expected. To explore this prediction, Coyne and Orr (1989) used an extensive literature survey of 119 *Drosophila* species pairs where data were available for at least one measure of reproductive isolation and information was available to estimate genetic distances. From this survey, they found strong support for a positive correlation between reproductive isolation and genetic distance (Coyne and Orr 1989, 1997). This research initiative was extended to other animal groups, and the general trend continued to hold (Mendelson 2003). Moyle et al. (2004) scoured the literature to apply the Coyne and Orr (1989) approach to published studies for plants and found three genera with suitable data. For one group, *Silene* L., they found a pattern of increasing reproductive isolation with greater genetic distances, consistent with the animal studies (Moyle et al. 2004). However, for the other two genera (*Glycine* Willd. and *Streptanthus* Nutt.), the pattern was not as straightforward (Moyle et al. 2004). They concluded that other factors may have contributed to the pattern of interfertility, such as insufficient time to develop effective reproductive barriers (*Streptanthus*), unusual crossing relationships of individual species (*Glycine falcata* Benth.), or perhaps genes of large effect causing the patterns of reproductive isolation (*Glycine*).

When considering whether a pair of plant species is likely to be crossed successfully, intuitively one might predict that species very different in appearance will be less likely to produce viable hybrid offspring than similar-looking species. For example, *Jaltomata* species exhibit striking floral diversity. One might expect the two species with rotate corollas, *Jaltomata sinuosa* (Miers) Mione and *J. antillana* (Krug & Urb.) D'Arcy, to be interfertile and similarly for the pair with tubular corollas, *J. aijana* Mione & S. Leiva and *J. incahuasina* Mione & S. Leiva (Figure 6.1). The notion is that traits related to floral divergence may have pleiotropic effects on development so that greater overall morphological divergence would indirectly lead to genetic incompatibilities and thereby result in reduced postmating prezygotic and intrinsic postzygotic reproductive isolation. Kostyun and Moyle (2017) tested this idea in an investigation of 10 species of *Jaltomata*. They did not find support for floral divergence being a strong predictor of the strength of reproductive isolation. In contrast, a significant correlation between genetic distance and intrinsic postzygotic reproductive isolation was detected, providing additional support for this general relationship.

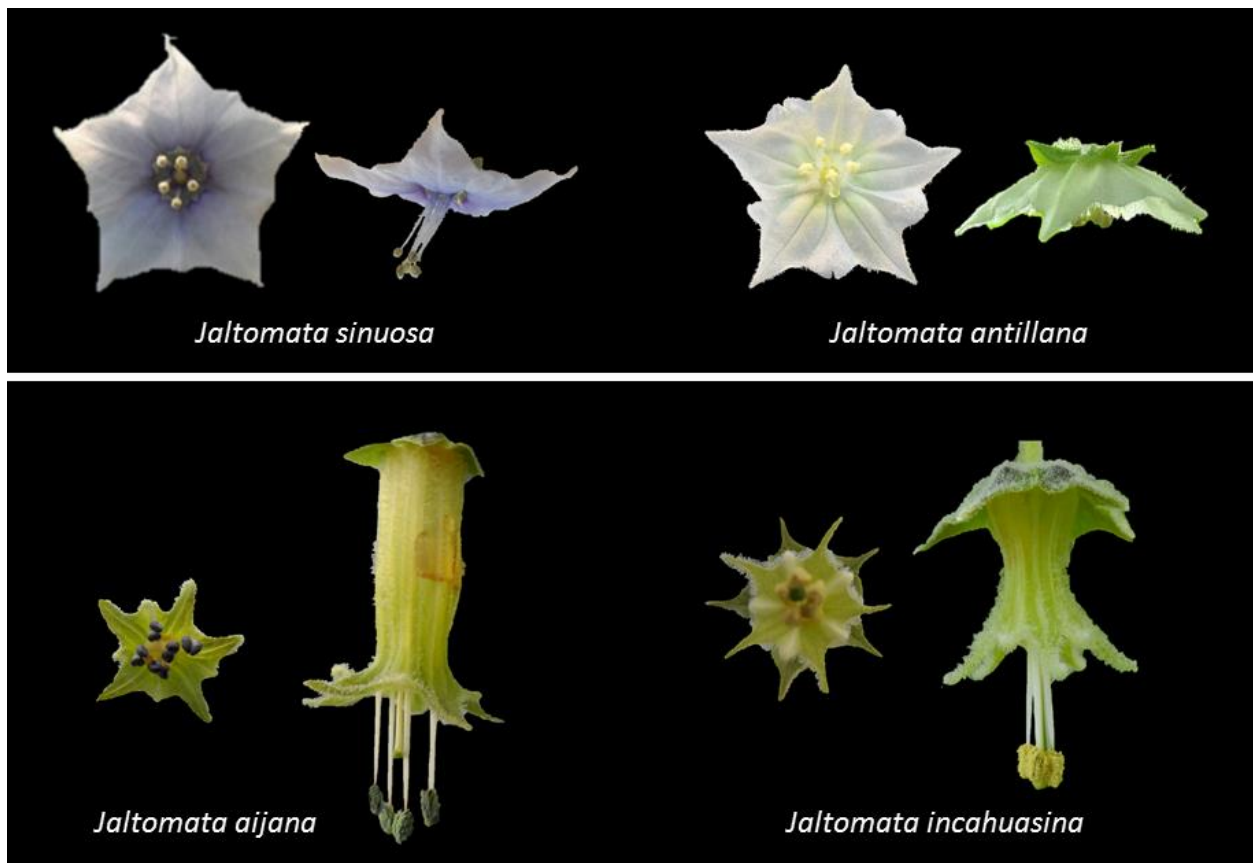


Fig. 6.1: Floral diversity among four *Jaltomata* species with two species exhibiting rotate corollas and two species with tubular corollas. (Photographs courtesy Jamie Kostyun)

Animal studies of reproductive isolation have shown that prezygotic isolating mechanisms are often stronger than postzygotic isolating mechanisms and evolve sooner between diverging species pairs (Coyne and Orr 1989, 1997; Mendelson 2003). The results from investigations of suites of individual reproductive barriers between plant species are consistent with findings from animals, with prezygotic isolating mechanisms, both before and after pollination, providing a much greater contribution than postzygotic isolating mechanisms to overall reproductive isolation (Rieseberg and Willis 2007; Lowry et al. 2008). For CWR, it is informative to consider prezygotic isolating mechanisms, such as ecogeographic separation or flowering time differences, to help to explain the diversification of these species. However, for plant breeding purposes, artificial crosses circumvent these prepollination prezygotic barriers to reproduction. Therefore, postpollination prezygotic barriers and intrinsic postzygotic barriers should be the primary focus for the successful utilization of CWR.

If the barriers to reproduction are the result of an accumulation of nuclear genetic changes acting in a Mendelian fashion, then the patterns of reproductive isolation would be predicted to be symmetrical regardless of which species serves as the sire and which serves as the dam (Tiffin et al. 2001). However, asymmetry in patterns of reproductive isolation is common in plants based on which species acts as the pollen or seed parent (Tiffin et al. 2001; Lowry et al. 2008; Plazas et al. 2016; Kostyun and Moyle 2017). In a survey of 14 genera, Tiffin et al. (2001) detected significant asymmetries in three life history stages: seed production, hybrid viability, and hybrid sterility. In another review, Lowry et al. (2008) found asymmetries were greatest for postmating barriers, notably hybrid seed set and viability. These asymmetries indicate factors that are likely to contribute to reproductive isolation, including gametophyte-sporophyte interactions (Pease et al. 2016), problems with endosperm development because of genomic incompatibilities causing defective seeds (Lafon-Placette and Köhler 2016), cytonuclear interactions that may lead to hybrid male sterility (Chen et al. 2017), and others (reviewed in Turelli and Moyle 2007).

An interesting example of a gametophyte-sporophyte interaction comes from an investigation of tomato (*Solanum lycopersicum*) and its relatives. In a study involving the wild species *Solanum pennellii* Correll, Pease et al. (2016) identified the molecular genetic basis of postmating prezygotic reproductive isolation. They examined the fascinating system of unilateral incompatibility, where *S. pennellii* (self-incompatible) pollen tubes can extend into *S. lycopersicum* (self-compatible) styles to fertilize the ovaries, whereas *S. lycopersicum* pollen grains are prematurely blocked in the styles of *S. pennellii*. Using an elegant experimental design, transcriptomes were obtained from a combination of tissues that allowed contrasting gene expression patterns to be used to identify candidate genes potentially responsible for the observed reproductive barriers (Pease et al. 2016).

6.8 Further Innovating the Gene Pool Concept for CWR

The focus on genetic relatedness of the Wiersema and León (2016) proposal is excellent and represents the best effort to date to use an understanding of evolutionary relationships to categorize CWR across many crop gene pools. However, with many thousands of potentially useful species to evaluate (Vincent et al. 2013; Dempewolf et al. 2014), a more widely applicable and standardized system of assessment would be extremely helpful, building on a quantitative understanding of the relationship between genetic relatedness and patterns of interfertility among

CWR. Three components stand out as required to accomplish the creation of this system. The first is to determine to the greatest extent possible evolutionary relationships among CWR, as has been discussed. The remaining components build on the understanding of the evolution of reproductive isolation in plants by developing two quantitative measures of the patterns of interfertility among CWR: observed and predicted gene pool indices.

The observed gene pool index can be constructed based on results from crossing studies involving CWR and crops. It is similar to estimating the level of reproductive isolation for studies of speciation by carefully examining all potential reproductive barriers and calculating their cumulative effect (Martin and Willis 2007; Lowry et al. 2008). However, the observed gene pool index places a primary focus on reproductive processes that reduce the level of interfertility related to fertilization success, seed viability, offspring vigor, and hybrid fertility. Separate values are needed based on which species is the female or male parent for each cross due to the aforementioned asymmetries in the action of reproductive barriers. The observed gene pool index should in sum reflect the probability of producing viable and fertile hybrid offspring from a select cross.

More challenging is transforming the observed gene pool index into a predicted gene pool index. As in the majority of cases crossing studies have not been comprehensively performed for all species related to all crops, the predictive index ends up in actuality being the foundation for developing an independent and reproducible assessment system.

The first step to develop the predicted gene pool index borrows from the approach of Coyne and Orr (1989, 1997) and Moyle et al. (2004) to use available results from the literature to survey relationships between genetic relatedness and patterns of interfertility among CWR and their crop species. Specifically, reproductive barriers related to postpollination prezygotic barriers and intrinsic postzygotic reproductive barriers are of interest. The objectives of the survey are to determine if a broad relationship exists between the observed gene pool index and measures of genetic relatedness (e.g., based on nucleotide substitutions, patristic distances, etc.). However, equally important is to bring attention to those cases that deviate from the general pattern and to consider various factors that may reduce interfertility in these cases, including gametophyte-sporophyte interactions disrupting fertilization, problems with endosperm development leading to inviable seeds, and cytonuclear interactions resulting in hybrid male sterility. There may be a quantitative relationship between interfertility and genetic distance, but an additional qualitative pattern may emerge owing to failed pollination, seeds dying, and hybrids being sterile. Therefore, in addition to obtaining quantitative estimates of genetic relatedness, it is important to identify characteristics of the CWR that may also determine patterns of interfertility, such as ploidy level, mating system, pollen size and style length, properties of the endosperm, and characteristics of the mitochondrial genome.

It may be necessary to carry out careful experimental crossing studies coupled with phylogenetic investigations for groups of CWR where a knowledge gap exists. In some cases, the number of populations of CWR to consider may be too numerous to be included in the labor-intensive efforts of crossing studies; therefore, a careful exemplar sample should reflect the important characteristics that may determine the pattern of interfertility. Further analyses providing results for various gene pools, and then testing the results across other plant groups, will be necessary to determine whether particular interfertility factors are highly important across most or all groups. With this knowledge, further studies on other groups can be more efficiently targeted.

A study of eggplant and its wild relatives (Plazas et al. 2016) provides an indication of the methods needed to develop an observed gene pool index. Crosses were carried out between 6 accessions of eggplant (*Solanum melongena*) and 19 accessions of 12 wild species, sampling from the three tiers of the gene pool (Plazas et al. 2016). Fruit set, number of seed per fruit, and germination success were recorded with *S. melongena* as both the female and male parent. Fruit set was as high as 48%, maximum seeds per fruit was 2.7, and in one case 92% germination was observed. From these values, the number of viable seeds produced from 100 crosses can be estimated. For example, the crosses between *S. melongena* and its progenitor *S. insanum* would result in 44 viable offspring (17.8% fruit set, 2.67 seeds/fruit, 92.2% germination) with eggplant as the female parent and 55 viable offspring (33.3% fruit set, 2.18 seeds/fruit, 75.8% germination) with eggplant as the male parent.

To develop a corollary predictive gene pool index for eggplant and its wild relatives, genetic data obtained by Acquadro et al. (2017) and the phylogenetic investigations by Levin et al. (2006), Vorontsova et al. (2013), and Aubriot et al. (2016) can provide the foundation. The taxa sampled by Vorontsova et al. (2013) included many species in the Plazas et al. (2016) study. Based on results from two nuclear genes and a chloroplast marker, Vorontsova et al. (2013) resolved clades that can be used as an estimate of patristic distance. The phylogeny indicates that *Solanum melongena* was nested with its putative progenitor, *S. insanum*, with *S. incanum* as the sister species to this clade. *Solanum linnaeanum* and *S. lichtensteinii* were the next closest relatives, whereas *S. violaceum*, *S. dasyphyllum* Schumacher & Thonn., *S. anguivi* Lam., *S. tomentosum*, *S. pyracanthos* Lam., *S. elaeagnifolium* Cav., and *S. torvum* Sw. were increasingly distantly related and were not part of the eggplant clade.

These phylogenetic relationships can be compared with the patterns of interfertility from the Plazas et al. (2016) crossing study, focusing on fruit set with the most complete data (percent fruit set for hybrids with eggplant as the female or male parent, respectively). A general pattern emerges where species more closely related to *S. melongena* had intermediate levels of fruit set: *S. insanum* (18, 33), *S. incanum* (18, 25), *S. linnaeanum* (9, 48), *S. lichtensteinii* (17, 18), *S. violaceum* (5, 0), *S. dasyphyllum* (24, 11), *S. anguivi* (15, 34), and *S. tomentosum* (12, 32). In contrast, more distantly related species had low levels of fruit set: *S. pyracanthos* (0, 5), *S. elaeagnifolium* (0, 0), and *S. torvum* (3, 0). It is intriguing to note that the species that deviated from the general pattern, *S. violaceum*, was a taxon that exhibited admixture and showed membership in lineages characteristic of three other species (Acquadro et al. 2017). Among the species more closely related to eggplant, no linear trend between genetic relatedness and fruit set was detected. This suggests that factors acting in a qualitative fashion are more likely to be determining the pattern of interfertility among eggplant and these CWR.

There is a vast literature on the systematics of crops and their wild relatives, and, in addition, there are decades of crossing studies involving CWR -- many more than could be considered here. The greatest gap at this point is the application of a quantitative approach to make a strong connection between evolutionary relationships and patterns of interfertility among CWR. Using published research to explore this relationship, looking for general trends, predictable patterns, and unusual special cases, is sure to provide insight and some surprises. Such a survey will help to identify gaps in knowledge that will hopefully inspire targeted further research. In combination, these investigations hold great promise to contribute to the development of the foundation understanding needed to anticipate with much greater accuracy how efficiently CWR can contribute to crop improvement.

References

- Acquadro A, Barchi L, Gramazio P, Portis E, Vilanova S, Comino C, Plazas M, Prohens J, Lanteri S (2017) Coding SNPs analysis highlights genetic relationships and evolution pattern in eggplant complexes. *PloS ONE* 12:e0180774
- Albert VA, Barbazuk WB, Der JP, Leebens-Mack J, Ma H, Palmer JD, Rounsley S et al. (2013) The Amborella genome and the evolution of flowering plants. *Science* 342:1241089
- Aubriot X, Paramjit S, Knapp S. (2016) Tropical Asian species show that the old world clade of ‘spiny solanums’ (*Solanum* subgenus *Leptostemonum* pro parte: Solanaceae) is not monophyletic. *Botanical Journal of the Linnean Society* 181:199-223.
- Austin DF (1978) The *Ipomoea batatas* complex-I. Taxonomy. *Bulletin of the Torrey Botanical Club* 1:114-29
- Austin DF (1997) Dissolution of *Ipomoea* series *Anisomerae* (Convolvulaceae). *Taxon* 28:359-361
- Baack E, Melo MC, Rieseberg LH, Ortiz-Barrientos D (2015) The origins of reproductive isolation in plants. *New Phytologist* 207:968-984.
- Barrett CF, Freudenstein JV (2011) An integrative approach to delimiting species in a rare but widespread mycoheterotrophic orchid. *Molecular Ecology* 20:2771-2786
- Bohs L (1991) Crossing studies in *Cyphomandra* (Solanaceae) and their systematic and evolutionary significance. *American Journal of Botany* 78:1683-1693
- Bredeson JV, Lyons JB, Prochnik, SE, Wu GA, Ha CM, Edsinger-Gonzales E, Grimwood J, Schmutz J, Rabbi IY, Egesi C, Nauluvula P (2016) Sequencing wild and cultivated cassava and related species reveals extensive interspecific hybridization and genetic diversity. *Nature Biotechnology* 34:562-570
- Brozynska M, Furtado A, Henry RJ (2016) Genomics of crop wild relatives: expanding the gene pool for crop improvement. *Plant Biotechnology Journal* 14:1070-1085
- Buddenhagen C, Lemmon AR, Lemmon EM, Bruhl J, Cappa J, Clement WL, Donoghue M et al. (2016) Anchored phylogenomics of angiosperms I: assessing the robustness of phylogenetic estimates. *bioRxiv*:086298
- Carstens BC, Pelletier TA, Reid NM, Satler JD (2013) How to fail at species delimitation. *Molecular Ecology* 22:369-4383
- Chen Z, Zhao N, Li S, Grover CE, Nie H, Wendel JF, Hua J, (2017) Plant mitochondrial genome evolution and cytoplasmic male sterility. *Critical Reviews in Plant Sciences* 36:55-69
- Comer JR, Zomlefer WB, Barrett CF, Stevenson DW, Heyduk K, Leebens-Mack JH (2016) Nuclear phylogenomics of the palm subfamily *Arecoideae* (Arecaceae). *Molecular Phylogenetics and Evolution* 97:32-42
- Coyne JA, Orr HA (1989) Patterns of speciation in *Drosophila*. *Evolution* 43:362-381
- Coyne JA, Orr HA (1997) Patterns of speciation in *Drosophila* revisited. *Evolution* 51:295
- Coyne JA, Orr HA (1998) The evolutionary genetics of speciation. *Philos Trans of the R Soc Lond B: Biological Sciences* 353:287-305
- De Queiroz K (2005) Ernst Mayr and the modern concept of species. *Proceedings of the National Academy of Sciences* 102:6600-6607

- De Queiroz K (2007) Species concepts and species delimitation. *Systematic Biology* 56:879-886
- Degnan JH, Rosenberg NA (2009) Gene tree discordance, phylogenetic inference and the multispecies coalescent. *Trends Ecol Evol* 24:332-340
- Dempewolf H, Eastwood RJ, Guarino L, Khoury CK, Müller JV, Toll J (2014) Adapting agriculture to climate change: a global initiative to collect, conserve, and use crop wild relatives. *Agroecol Sust Food* 38:369-377
- Diaz J, Schmiediche P, Austin DF (1996) Polygon of crossability between eleven species of *Ipomoea*: section *Batatas* (Convolvulaceae). *Euphytica* 88:189-200
- Donoghue MJ (1985) A critique of the biological species concept and recommendations for a phylogenetic alternative. *The Bryologist* 88:172-181
- Duarte JM, Wall PK, Edger PP, Landherr LL, Ma H, Pires PK, Leebens-Mack J (2010) Identification of shared single copy nuclear genes in *Arabidopsis*, *Populus*, *Vitis* and *Oryza* and their phylogenetic utility across various taxonomic levels. *BMC Evolutionary Biology* 10:61
- Duncan TM, Rausher MD (2013a) Evolution of the selfing syndrome in *Ipomoea*. *Frontiers in Plant Science* 4:301
- Duncan TM, Rausher MD (2013b) Morphological and genetic differentiation and reproductive isolation among closely related taxa in the *Ipomoea* series *Batatas*. *American Journal of Botany* 100:2183-2193
- Edwards SV (2009) Is a new and general theory of molecular systematics emerging? *Evolution* 63:1-19
- Ellstrand NC, Prentice HC, Hancock JF (1999) Gene flow and introgression from domesticated plants into their wild relatives. *Annual review of Ecology and Systematics* 30:539-563
- Eserman LA (2012) Taxonomy and crossing relationships in a small group of morning glories (*Ipomoea* section *Pharbitis*). Master thesis, Southeastern Louisiana University
- Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology* 14:2611-2620
- Faircloth BC, McCormack JE, Crawford NG, Harvey MG, Brumfield RT, Glenn TC (2012) Ultraconserved elements anchor thousands of genetic markers spanning multiple evolutionary timescales. *Systematic Biology* 61:717-726
- Falush D, Stephens M, Pritchard JK, (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* 164:1567-1587
- Fujita MK, Leaché AD, Burbrink FT, McGuire JA, Moritz C (2012) Coalescent-based species delimitation in an integrative taxonomy. *Trends Ecol Evol* 27:480-488
- Grant V (1981) Plant speciation. Columbia University, New York
- Grover CE, Gallagher JP, Jareczek JJ, Page JT, Udall JA, Gore MA, Wendel JF (2015) Reevaluating the phylogeny of allopolyploid *Gossypium* L. *Molecular phylogenetics and evolution* 92:45-52
- Hajjar R, Hodgkin T (2007) The use of wild relatives in crop improvement: a survey of developments over the last 20 years. *Euphytica* 156:1-13
- Harlan JR, De Wet MJM (1971) Toward a rational classification of cultivated plants. *Taxon* 20:509-517
- Hardigan MA, Bamberg J, Buell CR, Douches DS (2015) Taxonomy and genetic differentiation among wild and cultivated germplasm of *Solanum* sect. *Petota*. *Plant Genome* 8:1-16

- Heled J, Drummond AJ (2010) Bayesian inference of species trees from multilocus data. *Molecular Biology and Evolution* 27:570-580
- Hey J (2006) On the failure of modern species concepts. *Trends in Ecology & Evolution* 21:447-450
- Heyduk K, Stephens JD, Faircloth BC, Glenn TC (2016) Targeted DNA region re-sequencing. *Field Guidelines for Genetic Experimental Designs in High-Throughput Sequencing*, Cham Springer International Publishing, p 43-68
- Heyduk K, Trapnell DW, Barrett CF, Leebens-Mack J (2015) Phylogenomic analyses of species relationships in the genus *Sabal* (Arecaceae) using targeted sequence capture. *Biological Journal of the Linnean Society* 117:106-120
- Hirsch CD, Evans J, C Buell R, Hirsch CN (2014) Reduced representation approaches to interrogate genome diversity in large repetitive plant genomes. *Briefings in Functional Genomics* 13:257-267
- Hubisz MJ, Falush D, Stephens M, Pritchard JK (2009) Inferring weak population structure with the assistance of sample group information. *Molecular Ecology Resources* 9:1322-1332
- International Wheat Genome Sequencing Consortium (2014) A chromosome-based draft sequence of the hexaploid bread wheat (*Triticum aestivum*) genome. *Science* 345:1251788
- Jansky S (2006) Overcoming hybridization barriers in potato. *Plant breeding* 125:1-12
- Kay KM, Reeves PA, Olmstead RG, Schemske DW (2005) Rapid speciation and the evolution of hummingbird pollination in neotropical *Costus* subgenus *Costus* (Costaceae): evidence from nrDNA ITS and ETS sequences. *American Journal of Botany* 92:1899-1910
- Khoury CK, Heider B, Castañeda-Álvarez NP, Achicanoy HA, Sosa CC, Miller RE, Scotland RW, Wood JR, Rossel G, Eserman LA, Jarret RL (2015) Distributions, ex situ conservation priorities, and genetic resource potential of crop wild relatives of sweetpotato [*Ipomoea batatas* (L.) Lam., I. series *Batatas*]. *Frontiers in Plant Science* 6:251.
- Kostyun JL, Moyle LC (2017) Multiple strong postmating and intrinsic postzygotic reproductive barriers isolate florally diverse species of *Jaltomata* (Solanaceae). *Evolution* 6:1556-1571
- Lafon-Placette C, Köhler C (2016) Endosperm-based postzygotic hybridization barriers: developmental mechanisms and evolutionary drivers. *Molecular Ecology* 25:2620-2629
- Lemmon EM, Lemmon AR (2013) High-throughput genomic data in systematics and phylogenetics. *Annual Review of Ecology, Evolution, and Systematics* 44:99-121.
- Lerner HRL, Mindell DP (2005) Phylogeny of eagles, old world vultures, and other Accipitridae based on nuclear and mitochondrial DNA. *Molecular Phylogenetics and Evolution* 37:327-346
- Levin DA (1979) The nature of plant species. *Science* 204:381-384
- Levin RA, Myers NR, Bohs L (2006) Phylogenetic relationships among the "spiny solanums" (*Solanum* subgenus *Leptostemonum*, Solanaceae). *American Journal of Botany* 93:157-169
- Liu L, Wu S, Yu L (2015) Coalescent methods for estimating species trees from phylogenomic data. *Journal of Systematics and Evolution* 53:380-390
- Liu L, Yu L, Edwards SV (2010) A maximum pseudo-likelihood approach for estimating species trees under the coalescent model. *BMC Evolutionary Biology* 10:302
- Liu L, Yu L, Pearl DK, Edwards SV (2009) Estimating species phylogenies using coalescence times among sequences. *Systematic Biology* 58:468-77

- Lowry DB, Modliszewski JL, Wright KM, Wu CA, Willis JH (2008) The strength and genetic basis of reproductive isolating barriers in flowering plants. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* 363:3009-3021
- Luckow M (1995) Species concepts: assumptions, methods, and applications. *Systematic Botany* 20:589-605
- Mandel JR, Dikow RB, Funk VA, Masalia RR, Staton SE, Kozik A, Michelmore RW, Rieseberg LH, Burke JM (2014) A target enrichment method for gathering phylogenetic information from hundreds of loci: an example from the Compositae. *Applications in Plant Sciences* 2:1300085
- Martin NH, Willis JH (2007) Ecological divergence associated with mating system causes nearly complete reproductive isolation between sympatric *Mimulus* species. *Evolution* 61:68-82
- Maxted N, Ford-Lloyd BV, Jury S, Kell S, Scholten M (2006) Towards a definition of a crop wild relative. *Biodiversity and Conservation* 15:2673-2685
- Mascher M, Richmond TA, Gerhardt DJ, Himmelbach A, Clissold L, Sampath D, Ayling S et al. (2013) Barley whole exome capture: a tool for genomic research in the genus *Hordeum* and beyond. *The Plant Journal* 76:494-505
- Mayr E (1942) *Systematics and the Origin of Species*. Columbia University Press, New York
- McCormack JE, Hird SM, Zellmer AJ, Carstens BC, Brumfield R.T (2013) Applications of next-generation sequencing to phylogeography and phylogenetics. *Molecular Phylogenetics and Evolution* 66:526-538
- Mendelson TC (2003) Sexual isolation evolves faster than hybrid inviability in a diverse and sexually dimorphic genus of fish (Percidae:Etheostoma). *Evolution* 57:317-327
- Mirarab S, Warnow T (2015) ASTRAL-II: coalescent-based species tree estimation with many hundreds of taxa and thousands of genes. *Bioinformatics* 31:i44-i52
- Morrell PL, Buckler ES, Ross-Ibarra J (2012) Crop genomics: advances and applications. *Nature Reviews Genetics* 13:85-96
- Moyle LC, Olson MS, Tiffin P (2004) Patterns of reproductive isolation in three angiosperm genera. *Evolution* 58:1195-1208.
- Nordborg M (2001) Coalescent theory. *Handbook of Statistical Genetics*, John Wiley and Sons, Chichester
- O'Meara BC (2009) New heuristic methods for joint species delimitation and species tree inference. *Systematic Biology* 59:59-73
- Pease JB, Guerrero RF, Sherman NA, Hahn MW, Moyle LC (2016) Molecular mechanisms of postmating prezygotic reproductive isolation uncovered by transcriptome analysis. *Molecular Ecology* 25:2592-2608
- Petit RJ, Excoffier L (2009) Gene flow and species delimitation. *Trend Ecol Evol* 24:386-393
- Plazas M, Vilanova S, Gramazio P, Rodríguez-Burruezo A, Fita A, Herraiz FJ, Ranil R, Fonseka R, Niran L, Fonseka H, Kouassi B (2016) Interspecific hybridization between eggplant and wild relatives from different gene pools. *Journal of the American Society for Horticultural Science* 141:34-44
- Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155:945-959
- Prohens J, Gramazio P, Plazas M, Dempewolf H, Kilian B, Díez MJ, Fita A et al. (2017) Introgressomics: a new approach for using crop wild relatives in breeding for adaptation to climate change. *Euphytica* 213:158

- Raj A, Stephens M, Pritchard JK (2014) fastSTRUCTURE: variational inference of population structure in large SNP data sets. *Genetics* 197:573-589
- Rannala B, Yang Z (2017) Efficient Bayesian species tree inference under the multispecies coalescent. *Syst Biol* 66:823-842
- Rieseberg LH, Kim SC, Randell RA, Whitney KD, Gross BL, Lexer C, Clay K (2007) Hybridization and the colonization of novel habitats by annual sunflowers. *Genetica* 129:149-165
- Rieseberg LH, Willis JH, (2007) Plant speciation. *Science*:910-914
- Rosenberg NA, Nordborg M (2002) Genealogical trees, coalescent theory and the analysis of genetic polymorphisms. *Nature Reviews Genetics* 3: 380-390
- Small E (1984) Hybridization in the domesticated-weed-wild complex. In: Grant WF (ed) *Plant biosystematics*. Academic Press, Toronto, pp 195-210
- Smith SD, Baum DA (2007) Systematics of Iochrominae (Solanaceae): patterns in floral diversity and interspecific crossability. *Acta Horticulturae* 745:241-254
- Soltis DE, Gitzendanner MA, Stull G, Chester M, Chanderbali A, Chamala S, Jordon-Thaden I, Soltis PS, Schnable PS, Barbazuk WB (2013) The potential of genomics in plant systematics. *Taxon* 62:886-898
- Stebbins, GL (1950) *Variation and evolution in plants*. Columbia Biological Series, New York
- The Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408:796-815
- Tiffin P, Olson MS, Moyle LC (2001) Asymmetrical crossing barriers in angiosperms. *Proceedings of the Royal Society of London B: Biological Sciences* 268:861-867
- Turelli M, Moyle LC (2007) Asymmetric postmating isolation: Darwin's corollary to Haldane's rule. *Genetics* 176:1059-1088
- Verdcourt B 1963 Convolvulaceae In: Hubbard CE, Readhead E (eds) *Flora of Tropical East Africa*. Whitefriars Press, London, pp 1-161
- Vincent H, Wiersema J, Kell S, Fielder H, Dobbie S, Castañeda-Álvarez NP, Guarino L, Eastwood R, León B, Maxted N (2013) A prioritized crop wild relative inventory to help underpin global food security. *Biological Conservation* 167:265-275
- Vorontsova MS, Stern S, Bohs L, Knapp S (2013) African spiny *Solanum* (subgenus *Leptostemonum*, Solanaceae): a thorny phylogenetic tangle. *Botanical Journal of the Linnean Society* 173:176-193
- Wakeley J (2009) *Coalescent theory: an introduction*, Roberts and Company, Greenwood Village
- Wakeley J (2013) Coalescent theory has many new branches. *Theoretical Population Biology* 87:1
- Weese TL, Bohs L (2007) A three-gene phylogeny of the genus *Solanum* (Solanaceae). *Systematic Botany* 32:445-463
- Weitemier K, Straub SC, Cronn RC, Fishbein M, Schmickl R, McDonnell A, Liston A (2014) Hyb-Seq: combining target enrichment and genome skimming for plant phylogenomics. *Appl Plant Sci* 2:1400042

- Wiens JJ (2007) Species delimitation: new approaches for discovering diversity. *Systematic Biology* 56:875-878
- Wiens JJ, Servedio MR (2000) Species delimitation in systematics: inferring diagnostic differences between species. *Proc R Soc B* 267:631–636
- Wiersema JH, León B (2016) The GRIN taxonomy crop wild relative inventory. *Enhancing Crop Genepool Use: capturing wild relative and landrace diversity for crop improvement*, CAB International, p 453
- Wood JR, Carine MA, Harris D, Wilkin P, Williams B, Scotland RW (2015) *Ipomoea* (Convolvulaceae) in Bolivia. *Kew Bulletin* 70:31
- Yang Z, Rannala B (2010) Bayesian species delimitation using multilocus sequence data. *Proc Nat Acad Sci U S A* 107:9264–9269
- Zimin AV, Puiu D, Hall R, Kingan S, Salzberg SL (2017) The first near-complete assembly of the hexaploid bread wheat genome, *Triticum aestivum*. *Giga Science* 6:1-7