Report on the outcome of the AIMS project

Background

This project arose out the realization that hybrid chestnuts occur in naturally regenerated forests and in the orchards of chestnut growers. The hybrids in forests may occur due to sympatry with the native chinquapin species in the southern region of the previous native range of American chestnut. Hybrids in forests may also occur due to the naturalization and subsequent introgression of "intentional" hybrids, those made by chestnut growers hoping to improve the germplasm and those made by USDA, University, and nonprofit organization scientists to study the host range of chestnut blight and to introgress the naturally occurring resistance in most Chinese chestnuts in the American chestnut. The low species barriers in *Castanea* could also have resulted in "unintentional" hybrids, those resulting from outcrossing with species and hybrids in chestnut orchards to natural forest settings. These factors plus the loss or lack of records on the location of intentional hybrids could have resulted in admixed descendants in orchards and in natural settings. An additional complication is the difficulty of recognizing admixed trees, hybrid trees or even species atypical trees by morphology alone. In the context of this report, 'hybrid' means admixture consistent with F_1 hybrid. All other admixtures are simply reported as 'admixed'.

Aims of the AIMs project

1. Identify and develop a set of markers, each of which are polymorphic across all *Castanea* species, reproducible, accurate, scalable and platform independent.
2. Collect and genotype enough samples from putative "pure species" to detect admixture of species in any Castanea individual, at 5% or higher, for any combination of possible species.
3. Collect and genotype samples of naturally occurring American chestnut, other American chestnuts of paramount importance (e.g. Ellis), and chestnuts of interest to growers.
4. Optimize the approach to maintain accuracy, precision and scalability while at the same lowering the fully loaded cost per sample.

Results

The final dataset consisted of genotypes of 42 sequenced EST-SSR markers on each of 192 samples. The sample set included, as identified by the contributors, 42 *C. mollissima* (Chinese chestnut), six *C. henryi*, three *C. sequinii*, 22 *C. crenata* (Japanese chestnut), 18 *C. sativa* (European chestnut), 55 *C. dentata* (American chestnut), 13 *C. pumila* (Allegheny chinquapin), 33 *C. ozarkensis* (Ozark chinquapin), the chestnut cultivar hybrid 'Paragon' (*C. dentata/C. sativa*) and complex hybrid 'Luvall's Monster', of unknown ancestry. The samples included three sets of technical replicates and two sets of biological replicates.

The analysis method employed was Prichard's STRUCTURE[1], a Bayesian approach that is agnostic to human-assigned species labels. The method is not sensitive to the order of the data. This method tests the likelihood of a series of possible priors. The prior is how many groups there are (1 group, two groups, etc). The likelihood of each prior is tested, then compared with the others. The analysis detects the group composition of individual samples, given the prior. Thus, admixture estimates arise directly from the analysis without regard to what the humans think. The data were scored by repeatedly sequencing (~50x) through a simple sequence repeat (SSR) embedded in an expressed sequence to obtain accurate sequence and then counting the number of repeats.

The variation in technical and biological reps was due to missing data, not differences in allele calls. Missing data can generate "ghost admixture" estimates, the magnitude of which depend on the context of the entire dataset. In this data set, based on the replicate data, any admixture below 3% is likely to be spooky (i.e. unlikely to reappear again).

How the groups change as *K* goes from eight to six

Examining which grouping merges or splits at different values for the number of groups reveals how "robust" a group designation is. The groups shown (p1-6, attached pptx) are for *K* = 8, the current understanding of the number of putative species the data set includes. As *K* goes down (p7 pptx), the only groups to disappear are *C. seguinii*, which merges into the *C. henryi* group at *K* = 7, then both C. *sequinii* and *C. henryi* merge into admixtures of C. mollissima with either *C. ozarkensis* or *C. crenata*, at *K* = 6. The Evanno method (a method of selecting at which *K* value the data are most likely) chooses *K* = 6 [2]. This result is most likely driven by the small number of *C. henryi* and *C. sequinii* samples. Alternative interpretations are premature until the sample size of these two species is increased. Note that most of admixtures detected, including the Cape Elizabeth, Maine samples, do not change across these three groupings.

Some notable admixtures (by sample number)

Under presumed *C. mollissima*

11. Schmucki timber type: admixed with *C. crenata* and *C. seguinii*/*C. henryi*
25. Chestnut cultivar Heritage: admixed with *C. sativa.*

Under presumed *C. sequinii*

2. Tree possibly from Mo lut tsz, from China via S. Anagnostakis: Unadmixed *C. crenata*

Under presumed *C. crenata*

22. Tree thought to be possible *C. crenata/C. sativa* hybrid: unadmixed *C. mollissima*

Under presumed *C. sativa*

8 &9 These are identical: *C. sativa/C. crenata* hybrid
14. Berlin sativa: *C. sativa/C. dentata* hybrid

Under presumed *C. dentata*

10. Nursery stock tree: *C. sativa* admixed with *C. dentata*
11. Nursery stock tree: *C. sativa/C. crenata* hybrid
24. Naturally occurring tree: Evidence of admixture with *C. henryi/C. sequinii* (requires confirmation)
27. TACF breeding program tree: slight admixture with *C. mollissima*
28. TACF breeding program tree: Evidence of admixture with *C. ozarkensis* (requires confirmation)
29. TACF breeding program tree: admixed with *C. ozarkensis*.
38. TACF chapter breeding program tree: unadmixed *C. ozarkensis*
40. Cape Elizabeth, Maine: admixed with *C. sativa*
41. Cape Elizabeth, Maine: C. dentata/*C. sativa* hybrid
42. Cape Elizabeth, Maine: C. dentata/*C. sativa* hybrid
43. Cape Elizabeth, Maine: C. dentata/*C. sativa* hybrid
44. Cape Elizabeth, Maine: admixed with *C. sativa*
45. Cape Elizabeth, Maine: admixed with *C. sativa*
46. Cape Elizabeth, Maine: admixed with *C. sativa* and *C. crenata*
47. Cape Elizabeth, Maine: admixed with *C. sativa*
48. Cape Elizabeth, Maine: admixed with *C. crenata*
54. Naturally occurring progeny of native tree: admixed with *C. pumila*

Under presumed *C. pumila* or *C. ozarkensis*

1. Tree near Marshall, VA, presumed *C. pumila*: unadmixed *C. dentata*
13. Progeny of *C. pumila*/Johnson *C. ozarkensis: C. mollissima* admixed with *C. sativa and C. crenata*

Aims fulfilment

Aims of the AIMs project

1. Identify and develop a set of markers, each of which are polymorphic across all *Castanea* species, reproducible, accurate, scalable and platform independent.
2. Collect and genotype enough samples from putative "pure species" to detect admixture of species in any Castanea individual, at 5% or higher, for any combination of possible species.
3. Collect and genotype samples of naturally occurring American chestnut, other American chestnuts of paramount importance (e.g. Ellis), and a sample of chestnuts of interest to growers.
4. Optimize the approach to maintain accuracy, precision and scalability while at the same lowering the fully loaded cost per sample.

The first aim is fulfilled in all respects except the scalability. The method requires 100 samples to be cost-effective., given the next-gen sequencing approach. The second aim is fulfilled with respect to *C. mollissima* and *C. dentata*. The current collection of *C. crenata* and *C. sativa* are sufficient for the purpose of this analysis, but require 10 to 20 more unrelated trees of each species for the accurate estimate of ancestry involving three or more species. This aim is not fulfilled with respect to *C. henryi* and *C. sequinii*. This aim is also inadequately fulfilled for *C. pumila* and *C. ozarkensis*. Ten to fifteen more unrelated individuals of the Chinese chinquapins and *C. pumila* are needed. The third aim is not fulfilled in that not enough *C. dentata* could be included given the cost of the analysis.
The fourth aim is unfulfilled.

The next steps

Ron Revord at the University of Missouri and I at Notre Dame are funded to lead a participatory breeding program for the chestnut growers in the central United States. My part of this project will include the completion of aims two, three and four, above, followed by extensive genotyping of the germplasm available from growers. The latter activity will include generation of pedigrees as well as ascertainment of admixtures.

Conclusion

The results shown clearly show that unsuspected admixed *Castanea* occur in naturally regenerated forests, in the orchards of chestnut growers and in the orchards of breeding programs. Admixtures of American chestnuts and the native chinquapins are likely to be a long-standing natural result of range overlap. Some admixture with non-native *Castanea* may have preceded the appearance of ink disease and chestnut blight, at least in certain locations. Thus, consideration of what is "native," for the purpose of restoration, may be less important than consideration of ecological equivalence, at least under certain circumstances.

Submitted Monday, 18 May 2020

Jeanne Romero-Severson
University of Notre Dame

References1. Falush D, Stephens M, Pritchard J: **Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies**. *Genetics* 2003, **164**:1567 - 1587.
2. Earl DA, vonHoldt BM: **STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method**. *Conservation Genetics Resources* 2012, **4**(2):359-361.