Next-Generation Sampling: Pairing Genomics with Herbarium Specimens Provides Species-Level Signal in Solidago (Asteraceae)

Authors: Beck, James B., and Semple, John C.

Source: Applications in Plant Sciences, 3(6)

Published By: Botanical Society of America

URL: https://doi.org/10.3732/apps.1500014

Applications in Plant Sciences

Applications in Plant Sciences 2015 3(6): 1500014

Application Article

Next-generation sampling: Pairing genomics with herbarium specimens provides species-level signal in *Solidago* (Asteraceae)[1]

James B. Beck[2,3,5], and John C. Semple[4]

[2]Department of Biological Sciences, Wichita State University, 537 Hubbard Hall, Wichita, Kansas 67260 USA; [3]Botanical Research Institute of Texas, 1700 University Drive, Fort Worth, Texas 76107 USA; and [4]Department of Biology, University of Waterloo, Waterloo, Ontario NL2 3G1 Canada

- *Premise of the study:* The ability to conduct species delimitation and phylogeny reconstruction with genomic data sets obtained exclusively from herbarium specimens would rapidly enhance our knowledge of large, taxonomically contentious plant genera. In this study, the utility of genotyping by sequencing is assessed in the notoriously difficult genus *Solidago* (Asteraceae) by attempting to obtain an informative single-nucleotide polymorphism data set from a set of specimens collected between 1970 and 2010.
- *Methods:* Reduced representation libraries were prepared and Illumina-sequenced from 95 *Solidago* herbarium specimen DNAs, and resulting reads were processed with the nonreference Universal Network-Enabled Analysis Kit (UNEAK) pipeline. Multidimensional clustering was used to assess the correspondence between genetic groups and morphologically defined species.
- *Results:* Library construction and sequencing were successful in 93 of 95 samples. The UNEAK pipeline identified 8470 single-nucleotide polymorphisms, and a filtered data set was analyzed for each of three *Solidago* subsections. Although results varied, clustering identified genomic groups that often corresponded to currently recognized species or groups of closely related species.
- *Discussion:* These results suggest that genotyping by sequencing is broadly applicable to DNAs obtained from herbarium specimens. The data obtained and their biological signal suggest that pairing genomics with large-scale herbarium sampling is a promising strategy in species-rich plant groups.

Key words: genotyping by sequencing; herbarium specimens; next-generation sampling; *Solidago*; species delimitation.

Shallow genetic differentiation and sampling limitations combine to restrict our understanding of biodiversity and evolution in many species-rich plant groups. Although numerous strategies for obtaining powerful genomic data sets are emerging (reviewed in Lemmon and Lemmon, 2013; McCormack et al., 2013), we remain fundamentally restricted by our access to samples. Other than the adoption of silica gel as a tissue dessicant (Chase and Hills, 1991), samples needed for plant molecular systematics studies are obtained essentially as they were at the beginning of the DNA era (Palmer and Zamir, 1982; Doyle et al., 1985). Researchers still must field-collect the majority of material—a rewarding, but expensive and time-consuming task that often precludes taxonomically rigorous sampling of large groups (>100 species) during the course of a dissertation or 3-yr federally funded project. If we are serious about understanding biodiversity and evolution in species-rich clades, we therefore need a transformative approach to obtaining samples.

Extracting DNA from herbarium specimen tissue is an obvious solution, an idea dating from the earliest days of plant molecular systematics (Rogers and Bendich, 1985). This type of sampling is, however, still viewed by most as a way to supplement an otherwise field-collected data set. Although studies utilizing genomic data sets obtained from herbarium specimens are emerging, most involve the recovery of high-copy organelle and/or rDNA cistron regions (Straub et al., 2012; Stull et al., 2013; Besnard et al., 2014; Ripma et al., 2014), or are focused on adaptation within a single species (Vandepitte et al., 2014) or genome assembly of a single individual (Staats et al., 2013). Indeed, we are unaware of a study that has performed species delimitation or phylogeny reconstruction using a genome-wide data set obtained exclusively (or even largely) from herbarium material. Sampling exclusively from herbarium material would allow robust taxonomic and geographic sampling to be achieved rapidly, and if this sampling were performed under the guidance of expert taxonomists it would also ensure the strongest link between taxonomy and DNA. Sample sets obtained through this strategy, what we term "next-generation sampling," could then be subjected to next-generation genotyping and sequencing techniques, as these workflows are presumably applicable to the sheared DNAs obtained from museum specimens (Nachman, 2013; Stull et al., 2013; Burrell et al., 2015). These rich data

[1]Manuscript received 12 February 2015; revision accepted 28 April 2015.

The authors thank K. Hyma and S. Mitchell from the Genomic Diversity Facility (GDF, Biotechnology Resource Center, Cornell University) for crucial guidance, and P. Soltis, P. Wolf, and two anonymous reviewers for comments on an earlier manuscript version. Sequencing at the GDF was supported by the National Institutes of Health (1S10OD010693-01). This work was supported by the Wichita State University Department of Biological Sciences, by a University of Wisconsin–Milwaukee Research Growth Initiative Grant, and Natural Sciences and Engineering Discovery Grants to J.C.S.

[5]Author for correspondence: james.beck@wichita.edu

doi:10.3732/apps.1500014

sets would then allow for the biodiversity and phylogeny of species-rich groups to be rigorously established in a short time.

In this study, we explore the compatibility of this sampling strategy with a genomic single-nucleotide polymorphism (SNP) protocol in the goldenrods (*Solidago* L., Asteraceae), a genus of ca. 150 currently recognized taxa (Semple and Cook, 2006). Taxonomic uncertainty in *Solidago* is widely recognized (Fernald, 1950; Nesom, 1993), a problem stemming from a combination of low interspecific genetic divergence (Kress et al., 2005; Fazekas et al., 2008; Schilling et al., 2008; Fazekas et al., 2009; Peirson et al., 2013), polyploidy (Semple, 1992), and species richness. In this study, we attempt to obtain genomic SNP information with a genotyping by sequencing (GBS) approach in a set of 95 herbarium specimens representing three *Solidago* subsections. These approaches identify SNPs at thousands of points throughout the genome by generating and sequencing a reduced representation library (Narum et al., 2013). Obtaining a genomic data set that carries species-level signal in this difficult genus, using only herbarium material, would be a powerful demonstration of the link between genomics and the expansive incorporation of herbarium material.

METHODS

***Sampling and DNA extraction/assessment*—**Polyploidy adds additional complexity to GBS data collection and analysis, including reduced per-individual sequencing depth due to increased genome size, the complicating nature of additional gene copies for SNP identification, and the relative lack of sophisticated analytical tools for polyploid data sets. We therefore chose to include diploid samples only in this pilot study. Herbarium tissue was obtained from 95 specimens representing 23 species in three *Solidago* subsections: *Junceae* (Rydb.) G. L. Nesom, *Squarrosae* A. Gray, and *Triplinerviae* (Torr. & A. Gray) G. L. Nesom (Appendix 1). All material was sampled from specimens at the University of Waterloo Herbarium (WAT), now housed as a unit of the Université de Montréal Herbarium (MT). Diploid mitotic chromosome counts were available for 73 of the 95 specimens (Semple et al., 1981, 1984, 1993; Semple and Chmielewski, 1987; J. Semple, unpublished data), and all exhibited microsatellite profiles indicative of diploidy (i.e., no more than two alleles per locus [J. Beck, unpublished data]). These specimens represented both a wide age range (collected between 1970 and 2010) and a diverse array of drying regimes, from field-based forced air techniques (similar to Blanco et al., 2006) to standard drying cabinets utilizing light bulbs or heaters. Approximately 15 mg of tissue were subjected to a cetyltrimethylammonium bromide (CTAB) protocol modified for 96-well plates (Beck et al., 2012). This high-throughput protocol has a history of yielding DNA quantity/quality sufficient for sequencing and genotyping in both herbarium (Beck et al., 2012, 2014; Alexander et al., 2013) and silica-dried (Rothfels et al., 2013) tissue. Concentration was determined with a Qubit 2.0 fluorometer (Life Technologies, Carlsbad, California, USA), and fragment size distribution was visualized by running 100 ng of extract against a λ DNA-HindIII digest (New England Biolabs, Ipswich, Massachusetts, USA) on a 1% agarose gel.

***Library preparation, sequencing, and SNP calling*—**GBS library preparation (Elshire et al., 2011), sequencing, and SNP calling were performed at the Genomic Diversity Facility (GDF) at Cornell University's Biotechnology Resource Center. Trial libraries for one DNA were generated with three enzymes (*Ape*KI, *Eco*T221, *Pst*I). Visual inspection of Experion (Bio-Rad, Hercules, California, USA) traces revealed that all exhibited fragment sizes generally between 150–300 bp. *Ape*KI was excluded due to the larger fragment pool, and thus lower read depth per fragment, that would result from this five-base recognition enzyme. Of the two six-base recognition enzymes, *Eco*T221 was then chosen because it exhibited a slightly smaller fragment pool. Libraries prepared from the 95 samples and one blank negative control were sequenced in one lane on an Illumina HiSeq 2500 (Illumina, San Diego, California, USA). Given that a reference genome was not available, the Universal Network-Enabled Analysis Kit (UNEAK) nonreference pipeline (Lu et al., 2013) implemented in TASSEL version 3.0.160 (Glaubitz et al., 2014) was used for tag alignment and subsequent SNP calling. The barcode/sample keyfile and all pipeline XML configuration files are archived at the Dryad Digital Repository (http://dx.doi.org/10.5061/dryad.16pj5; Beck and Semple, 2015).

***Data filtering and multivariate clustering*—**TASSEL 4.3 was used to produce preliminary SNP data sets by implementing high and low levels of missing data filtering on the total SNP set identified by UNEAK. This filtering and all further analyses excluded four samples (noted in Appendix 1). Two subsection *Triplinerviae* individuals were placed in other subsections in preliminary analyses, which along with other unpublished results strongly suggests that these are mislabeled DNA samples. Also excluded were two subsection *Squarrosae* individuals exhibiting low sequence read levels (see below). High filtering recovered SNPs present in 70% of samples, whereas low filtering recovered SNPs present in 30% of samples. Both filtering levels enforced a >1% minor allele frequency. These preliminary data sets were subjected to the multidimensional clustering approach employed in the principal coordinates analysis with modal clustering (PCO-MC) workflow (Reeves and Richards, 2009). This approach identifies the most cohesive groups in a data set by simultaneously considering information on all informative axes of a principal coordinates analysis. These groups are ranked by a "stability value," which ranges from 0–100 and quantifies the relative density of the group in multidimensional space (Reeves and Richards, 2009). Many clustering approaches are available for the analysis of SNP data (Lawson and Falush, 2012), and we employed PCO-MC based on its computational efficiency and ability to objectively identify and rank clusters. Unlike popular methods such as STRUCTURE (Pritchard et al., 2000) and STRUCTURAMA (Huelsenbeck et al., 2011), PCO-MC does not incorporate a model of within-group Hardy–Weinberg equilibrium, an assumption that is unrealistic for sets of individuals sampled at different times across the range of a species. Instead, PCO-MC identifies groups of individuals with similar genotypes, as genotypic similarity is but one of many secondary criteria that can be used to identify lineages (Mallet, 1995; Hausdorf and Hennig, 2010) under the general lineage concept (de Queiroz, 2007). The correspondence between clusters identified by PCO-MC and morphologically defined species (morphospecies) at both filtering levels was assessed. Cluster/morphospecies correspondence at high and low filtering levels was qualitatively similar in subsection *Triplinerviae* and generally lower at high filtering in subsections *Squarrosae* and *Junceae*. Low-filtered data sets were therefore chosen for subsequent PCO-MC clustering.

RESULTS

***Sequencing success and SNP recovery*—**Extracted DNA concentrations ranged from 15–155 ng/μL (mean: 46.2 ± 23.6), and total DNA yield ranged from 1050–10,850 ng (mean: 3185.9 ± 1665.9) (Appendix 1). Only five samples exhibited DNA yields below the 1.5-μg minimum recommended by the GDF. Gel electrophoresis indicated that all extracts were at least partially sheared, exhibiting fragment sizes between >23 kb and <500 bp (Appendix S1). Each extract was given a qualitative score of DNA degradation (1 = mainly large fragments [>23 kb]; 2 = relatively even distribution of large to small fragments; 3 = mainly small fragments [<2 kb]) (Appendix 1, Appendix S1). These degradation scores were strongly related to specimen age, as all 21 group 1 DNAs (least degraded) were collected since 1992 (Appendix 1). Reduced representation library construction and Illumina sequencing yielded 230,232,173 (100 bp) reads. Of these, 197,917,774 were considered quality reads, exhibiting no N's in the first 72 bases and including both a full barcode and the expected remnant of the restriction cut site (Elshire et al., 2011). These quality reads were then collapsed into 18,947,823 identical sequence tags. The blank sample returned 7604 quality reads, which was 0.003% of the total quality reads and 0.04% of the mean quality reads (2,076,237) per nonblank sample. Two samples were designated as failures by the GDF based on a quality read number <10% of this mean. Overall, quality read number per sample was significantly lower in older specimens ($r^2 = 0.27$, $P = 6.8 \times 10^{-8}$; Fig. 1A).

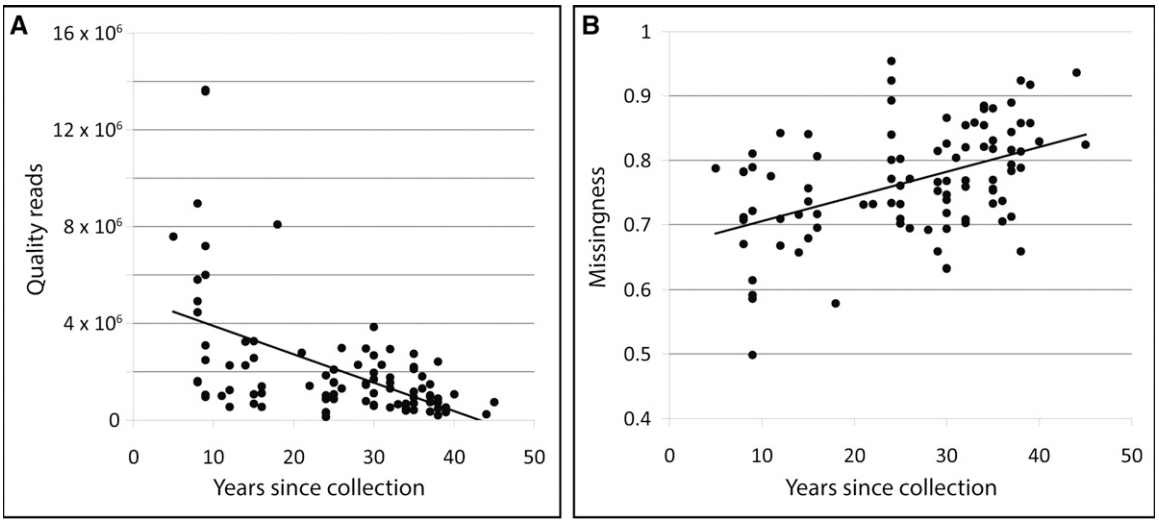

Fig. 1. Effect of *Solidago* specimen age on data quantity/quality. (A) Relationship between specimen age and the number of quality reads obtained for the 95 analyzed samples ($r^2 = 0.27$, $P = 6.8 \times 10^{-8}$). (B) Relationship between specimen age and the percentage of the 8470 unfiltered SNPs missing for the 95 analyzed samples ($r^2 = 0.22$, $P = 9.2 \times 10^{-7}$).

While still significant, the relationship between age and read number was less pronounced in specimens >10 yr old (r^2 = 0.080, P = 0.011). There were significant differences between the three DNA degradation categories [one-way ANOVA: $F(2,92) = 18.44$, $P < 0.0001$], with category 1 exhibiting more quality reads than categories 2 and 3 (Tukey honestly significant difference [HSD] test). The UNEAK pipeline identified 8470 unfiltered SNPs that were present in at least 10 of the 96 samples (blank included). Missingness, or the percentage of these SNPs exhibiting missing data in a given sample, was significantly higher in older specimens ($r^2 = 0.22$, $P = 9.2 \times 10^{-7}$) (Fig. 1B). There were again significant differences between the three DNA degradation categories [$F(2,92) = 20.44$, $P < 0.0001$], with category 1 DNAs exhibiting reduced missingness relative to category 2, which in turn exhibited reduced missingness relative to category 3 (Tukey HSD). Filtering to recover SNPs present in at least 70% of samples resulted in individual data sets of 547 (subsect. *Junceae*), 185 (subsect. *Squarrosae*), and 359 (subsect. *Triplinerviae*) SNPs. Filtering to recover SNPs present in at least 30% of samples resulted in individual data sets of 1633 (subsect. *Junceae*), 1447 (subsect. *Squarrosae*), and 2168 (subsect. *Triplinerviae*) SNPs. Original read data (FASTQ) have been deposited in the National Center for Biotechnology Information (NCBI) Sequence Read Archive (http://www.ncbi.nlm.nih.gov/sra) under BioProject ID PRJNA284163, and filtered subsection-specific HapMap matrices are archived at the Dryad Digital Repository (http://dx.doi.org/10.5061/dryad.16pj5; Beck and Semple, 2015).

Multivariate clustering—Correspondence between genetic groups identified by PCO-MC multidimensional clustering and morphospecies was strong in subsection *Junceae* (Fig. 2A). The five most highly ranked, and thus most cohesive in multivariate space, genetic clusters corresponded either to single morphospecies or groups of morphospecies. This result is particularly striking for the widespread species *S. missouriensis* Nutt. and *S. juncea* Aiton. In each case, samples from disparate portions of the morphospecies' range (*S. missouriensis* range shown in Fig. 2D) were identified as belonging to a significant genetic cluster. Also notable is the single incidence of a genetic cluster not corresponding to an entire morphospecies or group of morphospecies. PCO-MC identified a highly ranked cluster comprising all three TN specimens of the rare, strongly disjunct *S. gattingeri* Chapm. ex A. Gray, while the two samples from MO were not placed in this or any other cluster. This suggests that *S. gattingeri* comprises two morphologically cryptic species separated by the Mississippi Embayment (Fig. 2D), a hypothesis that is supported by multivariate morphological analyses (J. Semple, unpublished data). Correspondence between genetic clusters and morphospecies was also strong in subsection *Triplinerviae* (Fig. 2B). The six most cohesive clusters corresponded to single morphospecies (*S. gigantea* Aiton, *S. tortifolia* Elliott, and *S. elongata* Nutt.) or groups of morphospecies. While the four TX specimens of *S. juliae* G. L. Nesom composed a single cluster, the two AZ *S. juliae* specimens were not placed in this group. This again suggests the presence of two geographically disjunct species (Fig. 2D). The remaining samples, representing *S. altissima* L., *S. canadensis* L., *S. lepida* DC., and *S. brendiae* Semple, composed a single cluster. These species can at times be difficult to distinguish (Semple et al., 2013, 2015), and their lack of genetic distinctiveness is not unexpected. Although correspondence was not as strong in subsection *Squarrosae*, multiple highly ranked clusters corresponded to single morphospecies or putatively closely related morphospecies pairs (Fig. 2C). The most highly ranked cluster comprised all individuals of *S. pallida* (Porter) Rydb. and *S. rigidiuscula* (Torr. & A. Gray) Porter, two morphologically similar species that were until recently both part of the *S. speciosa* s.l. complex (Semple et al., 2012). All individuals of *S. erecta* Banks ex Pursh, another taxon historically placed in the *S. speciosa* complex, formed the next most highly ranked cluster with those of *S. speciosa* Nutt. itself. The third-ranked cluster comprised *S. puberula* Nutt. and two of three *S. pulverulenta* Nutt. individuals, two species that until recently were considered northern and southern subspecies of *S. puberula* s.l. (Semple and Cook, 2006; Fig. 2D). Finally, all three individuals of *S. squarrosa* Muhl. formed the fifth most highly ranked cluster.

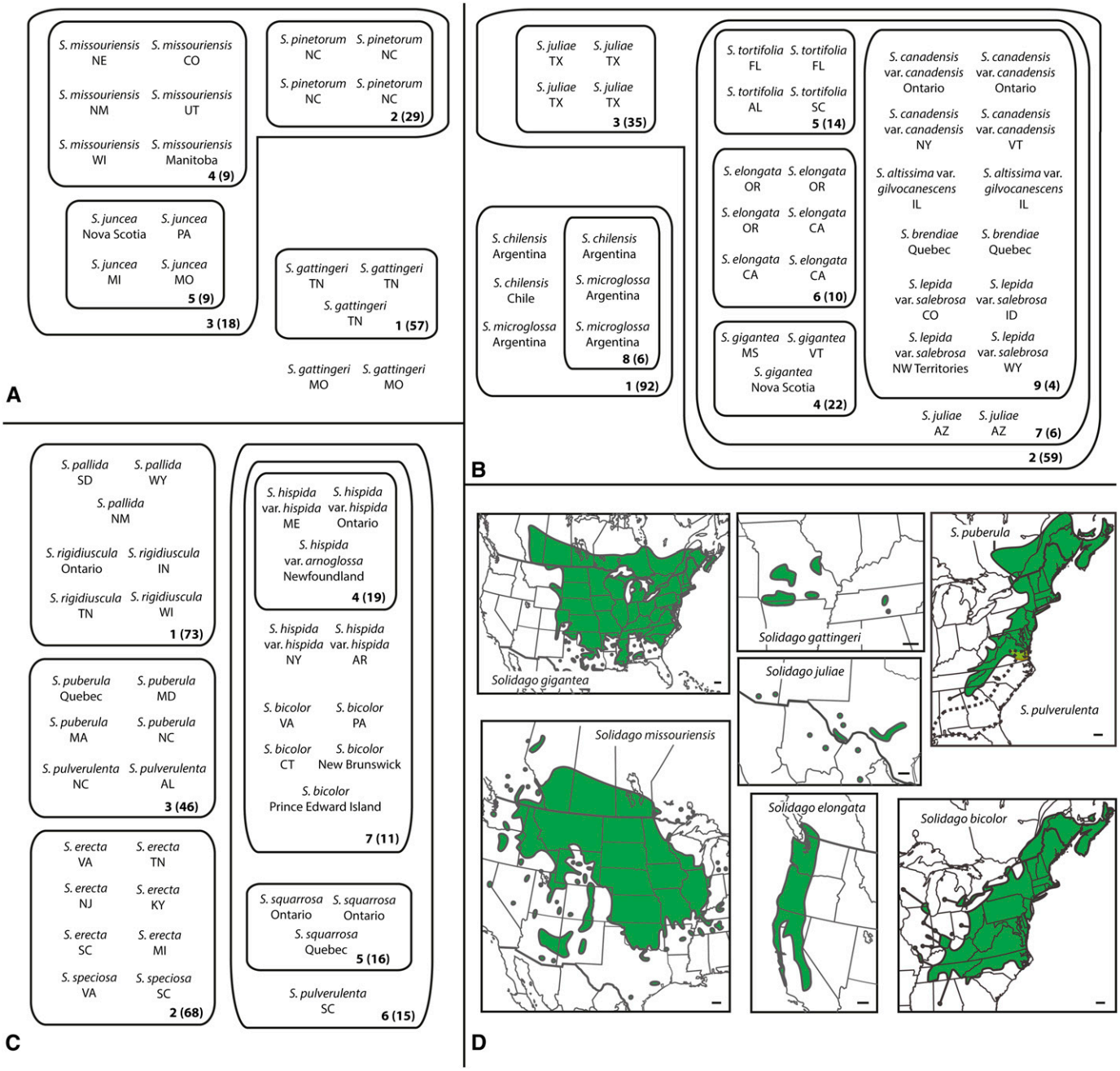

Fig. 2. Multidimensional clustering (PCO-MC) of GBS data for three *Solidago* subsections. (A) Graphical representation of the five most highly ranked, statistically significant clusters recovered for subsection *Junceae*. The rank of each cluster by stability (see Methods) and this value (in parentheses) appear at the bottom right of each cluster. Locality information for each specimen refers to the collection locality in Appendix 1. (B) Results for subsection *Triplinerviae*. (C) Results for subsection *Squarrosae*. (D) Range maps for select species (scale bars = 100 km).

DISCUSSION

We were able to routinely attain data at >1700 SNPs in a set of herbarium specimens representing 23 species obtained by numerous collectors over a 40-yr time span, and these data carried clear biological signal. Of the 20 strongest clusters identified by PCO-MC, seven comprised all individuals of a single species, two comprised clear geographic subsets of a single species, and three comprised all individuals of potentially sister species. This signal is particularly encouraging given the extremely low sequence divergence among goldenrod species. Schilling et al. (2008) observed <1% sequence divergence among *Solidago* species at the often highly variable internal transcribed spacer (ITS) region of the nuclear rDNA cistron. Among the eight groups examined in Kress et al. (2005), *Solidago* harbored the lowest level of diversity at 10 highly variable plastid loci, exhibiting no substitutions at the putatively universal barcoding region *psbA-trnH*. Fazekas et al. (2008, 2009) examined nine potential barcoding regions in 32 genera and commented that *Solidago* was one of the two most "intractable" genera. It

should also be noted that the inability of these data to recover clusters corresponding to all morphospecies may simply reflect biological reality, as it is unlikely that all currently recognized goldenrod species correspond to genetically cohesive groups (Semple and Cook, 2006). Taken together, these results indicate that the pairing of GBS with next-generation sampling holds considerable promise for species delimitation in large groups.

***Recommendations*—**We were able to consistently recover DNA of sufficient quantity/quality for library construction with a standard CTAB extraction protocol modified for 96-well plates, and the inexpensive and high-throughput nature of this approach pairs well with the large sample sizes we propose. Although specimen age did negatively affect both the number of quality reads and the amount of missing data per sample (Fig. 1), this effect was less pronounced for specimens >10 yr old. This suggests that much of this detrimental effect occurs at the time of collection (drying technique or length of time the sample was held before drying) or during the early years of curation, an insight consistent with studies that have explicitly evaluated the timing of DNA damage (Staats et al., 2011) and shearing (Adams and Sharma, 2010; Neubig et al., 2014) in herbarium material. Sampling could perhaps then be focused on relatively recent specimens if sufficient material is available. Specimen preparation practices and storage conditions have also been shown to exert a strong effect on DNA quality (Ribeiro and Lovato, 2007; Särkinen et al., 2012; Lander et al., 2013; Neubig et al., 2014), and sampling from air-dried material stored in humidity/temperature-controlled facilities should be favored. Following DNA extraction, our data suggest that a qualitative gel-based assessment of DNA degradation can be a strong predictor of downstream success. Regardless, future studies will need to evaluate the timing and degree of herbarium DNA degradation in a range of plant groups, as this process has been shown to proceed at varying rates in different taxa (Neubig et al., 2014).

Future studies could greatly enhance SNP discovery by beginning with low-coverage sequencing of one target species. The reference-aided GBS Discovery pipeline is robust to higher levels of divergence during locus identification and often identifies more SNPs, particularly in diverse data sets. Even a highly fragmentary assembly greatly improves SNP discovery, because short (64 bp) GBS reads can be matched to very small contigs. Genome size should also be considered. A recently examined diploid *Solidago* species exhibited a 1C-value = 1.02 pg (Kubešová et al., 2010), which is considered a relatively small angiosperm holoploid genome size (Leitch and Leitch, 2013). Genome size estimates across the group of interest should be considered during project design, particularly in the choice of restriction enzyme (Elshire et al., 2011). If funds permit, additional sequencing can be performed to reduce missingness in large-genome taxa (Chen et al., 2013). We also recommend the inclusion of multiple replicate samples to assess the background error rate. This is expected to be particularly important at the low read depths likely to be encountered in studies incorporating large numbers of specimens with varying DNA quality. Regarding analysis, a clear limitation of the cluster analysis of GBS data are the inability to reconstruct the pattern/timing of divergence among inferred lineages (Carstens et al., 2013), and fully leveraging these data for species delimitation and phylogeny reconstruction will require analytical tools that allow species trees to be inferred with the short read data obtained with GBS methods (Cariou et al., 2013; Hipp et al., 2014). These tools will no doubt soon be available (Leaché et al., 2014), as will increasingly longer read lengths of reduced representation libraries. These considerations notwithstanding, we feel strongly that pairing herbarium collections with GBS and other increasingly accessible genomic workflows (Straub et al., 2012; Stull et al., 2013; Weitemier et al., 2014) should be a top priority in plant systematics. Besides allowing for rapid and economical sampling of large groups, next-generation sampling allows specimen selection to be performed in collaboration with group experts. Genomic data sets spanning both species' ranges and intra/interspecific morphological variation can then be used to rigorously test a wide range of hypotheses, thanks to the synergy between big data and big sampling.

LITERATURE CITED

Adams, R. P., and L. N. Sharma. 2010. DNA from herbarium specimens: I. Correlation of DNA size with specimen age. *Phytologia* 92: 346–353.

Alexander, P. J., M. D. Windham, J. B. Beck, I. A. Al-Shehbaz, L. Allphin, and C. Donovan Bailey. 2013. Molecular phylogenetics and taxonomy of the genus *Boechera* and related genera (Brassicaceae: Boechereae). *Systematic Botany* 38: 192–209.

Beck, J. B., and J. C. Semple. 2015. Data from: Next-generation sampling: Pairing genomics with herbarium specimens provides species-level signal in *Solidago* (Asteraceae). Dryad Digital Repository. http://dx.doi.org/10.5061/dryad.16pj5.

Beck, J. B., P. J. Alexander, L. Allphin, I. A. Al-Shehbaz, C. Rushworth, C. D. Bailey, and M. D. Windham. 2012. Does hybridization drive the transition to asexuality in diploid *Boechera? Evolution* 66: 985–995.

Beck, J. B., J. C. Semple, J. M. Brull, S. L. Lance, M. M. Phillips, S. B. Hoot, and G. A. Meyer. 2014. Genus-wide microsatellite primers for the goldenrods (*Solidago*; Asteraceae). *Applications in Plant Sciences* 2: 1300093.

Besnard, G., P. A. Christin, P. J. G. Malé, E. Lhuillier, C. Lauzeral, E. Coissac, and M. S. Vorontsova. 2014. From museums to genomics: Old herbarium specimens shed light on a C_3 to C_4 transition. *Journal of Experimental Botany* 65: 6711–6721.

Blanco, M. A., W. M. Whitten, D. S. Penneys, N. H. Williams, K. M. Neubig, and L. Endara. 2006. A simple and safe method for rapid drying of plant specimens using forced-air space heaters. *Selbyana* 27: 83–87.

Burrell, A. S., T. R. Disotell, and C. M. Bergey. 2015. The use of museum specimens with high-throughput DNA sequencers. *Journal of Human Evolution* 79: 35–44.

Cariou, M., L. Duret, and S. Charlat. 2013. Is RAD-seq suitable for phylogenetic inference? An in silico assessment and optimization. *Ecology and Evolution* 3: 846–852.

Carstens, B. C., T. A. Pelletier, N. M. Reid, and J. D. Satler. 2013. How to fail at species delimitation. *Molecular Ecology* 22: 4369–4383.

Chase, M. W., and H. H. Hills. 1991. Silica gel: An ideal material for field preservation of leaf samples for DNA studies. *Taxon* 40: 215–220.

Chen, C., S. E. Mitchell, R. J. Elshire, E. S. Buckler, and Y. A. El-Kassaby. 2013. Mining conifers' mega-genome using rapid and efficient multiplexed high-throughput genotyping-by-sequencing (GBS) SNP discovery platform. *Tree Genetics & Genomes* 9: 1537–1544.

De Queiroz, K. 2007. Species concepts and species delimitation. *Systematic Biology* 56: 879–886.

Doyle, J. J., D. E. Soltis, and P. S. Soltis. 1985. An intergeneric hybrid in the Saxifragaceae: Evidence from ribosomal RNA genes. *American Journal of Botany* 72: 1388–1391.

Elshire, R. J., J. C. Glaubitz, Q. Sun, J. A. Poland, K. Kawamoto, E. S. Buckler, and S. E. Mitchell. 2011. A robust, simple Genotyping-by-Sequencing (GBS) approach for high diversity species. *PLoS ONE* 6: e19379.

Fazekas, A. J., K. S. Burgess, P. R. Kesanakurti, S. W. Graham, S. G. Newmaster, B. C. Husband, D. M. Percy, et al. 2008. Multiple multilocus DNA barcodes from the plastid genome discriminate plant species equally well. *PLoS ONE* 3: e2802.

Fazekas, A. J., P. R. Kesanakurti, K. S. Burgess, D. M. Percy, S. W. Graham, S. C. H. Barrett, S. G. Newmaster, et al. 2009. Are plant species inherently harder to discriminate than animal species using DNA barcoding markers? *Molecular Ecology Resources* 9: 130–139.

Fernald, M. L. 1950. Gray's manual of botany. Dioscorides Press, Portland, Oregon, USA.

Glaubitz, J. C., T. M. Casstevens, F. Lu, J. Harriman, R. J. Elshire, Q. Sun, and E. S. Buckler. 2014. TASSEL-GBS: A high capacity genotyping by sequencing analysis pipeline. *PLoS ONE* 9: e90346.

Hausdorf, B., and C. Hennig. 2010. Species delimitation using dominant and codominant multilocus markers. *Systematic Biology* 59: 491–503.

Hipp, A. L., D. A. R. Eaton, J. Cavender-Bares, E. Fitzek, R. Nipper, and P. S. Manos. 2014. A framework phylogeny of the American oak clade based on sequenced RAD data. *PLoS ONE* 9: e93975.

Huelsenbeck, J. P., P. Andolfatto, and E. T. Huelsenbeck. 2011. Structurama: Bayesian inference of population structure. *Evolutionary Bioinformatics Online* 7: 55–59.

Kress, J. W., K. J. Wurdack, E. A. Zimmer, L. A. Weight, and D. H. Janzen. 2005. Use of DNA barcodes to identify flowering plants. *Proceedings of the National Academy of Sciences, USA* 102: 8369–8383.

Kubešová, M., L. Moravcová, J. Suda, V. Jarošík, and P. Pyšek. 2010. Naturalized plants have smaller genomes than their non-invading relatives: A flow cytometric analysis of the Czech alien flora. *Preslia* 82: 81–96.

Lander, T. A., B. Dadonaite, and A. K. Monro. 2013. Microwave drying of plant material for herbarium specimens and genetic analysis. *Taxon* 62: 790–797.

Lawson, D. J., and D. Falush. 2012. Population identification using genetic data. *Annual Review of Genomics and Human Genetics* 13: 337–361.

Leaché, A. D., M. K. Fujita, V. N. Minin, and R. R. Bouckaert. 2014. Species delimitation using genome-wide SNP data. *Systematic Biology* 63: 534–542.

Leitch, I. J., and A. R. Leitch. 2013. Genome size diversity and evolution in land plants. *In* J. Greilhuber, J. Dolezel, and J. F. Wendel [eds.], Plant genome diversity, Vol. 2: Physical structure, behavior and evolution of plant genomes, 307–322. Springer, Vienna, Austria.

Lemmon, E. M., and A. R. Lemmon. 2013. High-throughput genomic data in systematics and phylogenetics. *Annual Review of Ecology, Evolution, and Systematics* 44: 99–121.

Lu, F., A. E. Lipka, J. Glaubitz, R. Elshire, J. H. Cherney, M. D. Casler, E. S. Buckler, and D. E. Costich. 2013. Switchgrass genomic diversity, ploidy, and evolution: Novel insights from a network-based SNP discovery protocol. *PLoS Genetics* 9: e1003215.

Mallet, J. 1995. A species definition for the modern synthesis. *Trends in Ecology & Evolution* 10: 294–299.

McCormack, J. E., S. M. Hird, A. J. Zellmer, B. C. Carstens, and R. T. Brumfield. 2013. Applications of next-generation sequencing to phylogeography and phylogenetics. *Molecular Phylogenetics and Evolution* 66: 526–538.

Nachman, M. W. 2013. Genomics and museum specimens. *Molecular Ecology* 22: 5966–5968.

Narum, S. R., C. A. Buerkle, J. W. Davey, M. R. Miller, and P. A. Hohenlohe. 2013. Genotyping-by-sequencing in ecological and conservation genomics. *Molecular Ecology* 22: 2841–2847.

Nesom, G. 1993. Taxonomic infrastructure of *Solidago* and *Oligoneuron* (Asteraceae: Astereae) and observations on their phylogenetic position. *Phytologia* 75: 1–44.

Neubig, K. M., W. M. Whitten, J. R. Abbott, S. Elliott, D. E. Soltis, and P. S. Soltis. 2014. Variables affecting DNA preservation in archival plant specimens. *In* W. L. Applequist and L. M. Campbell [eds.], DNA banking for the 21st century: Proceedings of the U.S. Workshop on DNA Banking, 81–112. William L. Brown Center, Missouri Botanical Garden, St. Louis, Missouri, USA.

Palmer, J. D., and D. Zamir. 1982. Chloroplast DNA evolution and phylogenetic relationships in *Lycopersicon. Proceedings of the National Academy of Sciences, USA* 79: 5006–5010.

Peirson, J. A., C. W. Dick, and A. A. Reznicek. 2013. Phylogeography and polyploid evolution of North American goldenrods (*Solidago* subsect. *Humiles*, Asteraceae). *Journal of Biogeography* 40: 1887–1898.

Pritchard, J. K., M. Stephens, and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. *Genetics* 155: 945–959.

Reeves, P. A., and C. M. Richards. 2009. Accurate inference of subtle population structure (and other genetic discontinuities) using principal coordinates. *PLoS ONE* 4: e4269.

Ribeiro, R. A., and M. B. Lovato. 2007. Comparative analysis of different DNA extraction protocols in fresh and herbarium specimens of the genus *Dalbergia. Genetics and Molecular Research* 6: 173–187.

Ripma, L. A., M. G. Simpson, and K. Hasenstab-Lehman. 2014. Geneious! Simplified genome skimming methods for phylogenetic systematic studies: A case study in *Oreocarya* (Boraginaceae). *Applications in Plant Sciences* 2: 1400062.

Rogers, S. O., and A. J. Bendich. 1985. Extraction of DNA from milligram amounts of fresh, herbarium and mummified plant tissues. *Plant Molecular Biology* 5: 69–76.

Rothfels, C. J., M. D. Windham, and K. M. Pryer. 2013. A plastid phylogeny of the cosmopolitan fern family Cystopteridaceae (Polypodiopsida). *Systematic Botany* 38: 295–306.

Särkinen, T., M. Staats, J. E. Richardson, R. S. Cowan, and F. T. Bakker. 2012. How to open the treasure chest? Optimising DNA extraction from herbarium specimens. *PLoS ONE* 7: e43808.

Schilling, E. E., J. B. Beck, P. J. Calie, and R. L. Small. 2008. Molecular analysis of *Solidaster* cv. *Lemore*, a hybrid goldenrod (Asteraceae). *Journal of the Botanical Research Institute of Texas* 2: 7–18.

Semple, J. C. 1992. A geographic summary of chromosome number reports for North American asters and goldenrods (Asteraceae: Astereae). *Annals of the Missouri Botanical Garden* 79: 95–109.

Semple, J. C., and J. G. Chmielewski. 1987. Chromosome number determinations in family Compositae, tribe Astereae. II. Additional counts. *Rhodora* 89: 319–325.

Semple, J. C., and R. E. Cook. 2006. *Solidago. In* Flora of North America Editorial Committee [eds.], Flora of North America, Vol. 20, Asteraceae Pt. II, 107–166. Oxford University Press, New York, New York, USA.

Semple, J. C., R. A. Brammall, and J. Chmielewski. 1981. Chromosome numbers of goldenrods, *Euthamia* and *Solidago* (Compositae-Astereae). *Canadian Journal of Botany* 59: 1167–1173.

Semple, J. C., G. S. Ringius, C. Leeder, and G. Morton. 1984. Chromosome numbers of goldenrods, *Euthamia* and *Solidago* (Compositae-Astereae). II. Additional counts with comments on cytogeography. *Brittonia* 36: 280–292.

Semple, J. C., J. Zhang, and C. Xiang. 1993. Chromosome number determinations in family Compositae, tribe Astereae. V. Eastern North American taxa. *Rhodora* 95: 234–253.

Semple, J. C., L. Tong, M. J. Oldham, and W. D. Bakowsky. 2012. *Solidago pallida* (Asteraceae: Astereae) new to Ontario and Canada. *Phytoneuron* 106: 1–5.

Semple, J. C., H. Faheemuddin, Y. A. Chong, M. K. Sorour, J. A. Hood, I. Khamis, Y. Ma, and K. Kornobis. 2013. A multivariate morphometric study of the *Solidago canadensis/S. lepida* complex of *Solidago* subsect. *Triplinerviae*. I. Northeastern taxa (Asteraceae: Astereae). *Phytoneuron* 58: 1–20.

Semple, J. C., H. Rahman, S. Bzovsky, M. K. Sorour, K. Kornobis, R. Lopez Lahitz, and L. Tong. 2015. A multivariate morphometric study of the *Solidago altissima* complex and *S. canadensis* (Asteraceae: Astereae). *Phytoneuron* 10: 1–31.

Staats, M., A. Cuenca, J. E. Richardson, R. Vrielnik-Van Ginkel, G. Petersen, O. Seberg, and F. T. Bakker. 2011. DNA damage in plant herbarium tissue. *PLoS ONE* 6: e28448.

Staats, M., R. H. Erkens, B. van de Vossenberg, J. J. Wieringa, K. Kraaijeveld, B. Stielow, J. Geml, et al. 2013. Genomic treasure troves: Complete genome sequencing of herbarium and insect museum specimens. *PLoS ONE* 8: e69189.

Straub, S. C. K., M. Parks, K. Weitemier, M. Fishbein, R. C. Cronn, and A. Liston. 2012. Navigating the tip of the genomic iceberg: Next-generation sequencing for plant systematics. *American Journal of Botany* 99: 349–364.

Stull, G. W., M. J. Moore, V. S. Mandala, N. A. Douglas, H. Kates, X. Qi, S. F. Brockington, et al. 2013. A targeted enrichment strategy for massively parallel sequencing of angiosperm plastid genomes. *Applications in Plant Sciences* 1: 1200497.

Vandepitte, K., T. Meyer, K. Helsen, K. Acker, I. Roldán-Ruiz, J. Mergeay, and O. Honnay. 2014. Rapid genetic adaptation precedes the spread of an exotic plant species. *Molecular Ecology* 23: 2157–2164.

Weitemier, K., S. C. K. Straub, R. C. Cronn, M. Fishbein, R. Schmickl, A. McDonnell, and A. Liston. 2014. Hyb-Seq: Combining target enrichment and genome skimming for plant phylogenomics. *Applications in Plant Sciences* 2: 1400042.

Appendix 1. Voucher information for *Solidago* individuals included in this study.

Species	Voucher specimen accession no.[a]	Collection year	Collection locality[b]	County[b]	DNA concentration	DNA yield	Gel image well (score)	Quality reads[c]	Missingness[d]
S. canadensis L. var. *canadensis*	Cook and Faulkenham C-14	1999	Ontario	Bruce Co.	38.7	2709	A01 (2)	1,393,599	0.717
S. canadensis var. *canadensis*	Semple and Brouillet 3667	1978	New York	Hamilton Co.	55.9	3913	C01 (2)	744,653	0.794
S. canadensis var. *canadensis*	Semple and Brouillet 3446	1978	Vermont	Washington Co.	60.8	4256	D01 (2)	1,488,711	0.712
S. canadensis var. *canadensis*	Semple & K. Shea 2416	1976	Ontario	Russell Co.	38.6	2702	E01 (3)	518,218	0.858
S. elongata Nutt.	Semple and Brouillet 7100	1983	Oregon	Hood Co.	33.5	2345	F01 (3)	527,280	0.854
S. elongata	Semple and Brouillet 7170	1983	Oregon	Lane Co.	41.3	2891	G01 (3)	1,301,975	0.769
S. elongata	Semple and Brouillet 7151A	1983	Oregon	Douglas Co.	56.4	3948	H01 (3)	1,303,240	0.759
S. elongata	Semple and Heard 8460	1986	California	Siskiyou Co.	62.4	4368	A02 (3)	793,643	0.815
S. elongata	Semple and Heard 8416	1986	California	Plumas Co.	57.9	4053	B02 (2)	2,959,297	0.659
S. elongata	Semple and Heard 8660	1986	California	Tulare Co.	89.1	6237	C02 (2)	1,532,168	0.753
S. altissima L. var. *gilvocanescens* (Rydb.) Semple	Semple and Brouillet 7367	1983	Illinois	Adams Co.	32.8	2296	D02 (2)	1,551,030	0.708
S. altissima var. *gilvocanescens*	Semple and Heard 8329	1985	Illinois	Johnson Co.	37.2	2604	E02 (2)	1,704,360	0.719
S. gigantea Aiton	Semple and Keir 4721	1980	Nova Scotia	Cumberland Co.	22.8	1596	F02 (2)	2,197,725	0.754
S. gigantea	Semple and Keir 4960	1980	Vermont	Windham Co.	42.2	2954	G02 (3)	2,113,381	0.756
S. gigantea	Semple and Suripto 10165	1991	Mississippi	Lowndes Co.	41.6	2496	H02 (2)	883,353	0.840
S. juliae G. L. Nesom	Morton and Venn NA16373	1985	Texas	Kendall Co.	47.8	3346	A03 (2)	1,952,178	0.739
S. juliae	Morton and Venn NA16370	1985	Texas	Kendall Co.	39.2	2744	B03 (2)	1,674,823	0.747
S. juliae	Nesom 7219	1989	Texas	Blanco Co.	64.4	4508	C03 (2)	1,313,653	0.772
S. juliae	Reeves R4521	1975	Arizona	Cochise Co.	54.7	3829	D03 (3)	1,056,658	0.829
S. juliae	Keil 18989	1985	Arizona	Santa Cruz Co.	71.8	5026	E03 (2)	3,845,527	0.694
S. juliae	Nesom 7213	1989	Texas	Real Co.	39.9	2793	F03 (2)	2,986,433	0.695
S. tortifolia Elliott[e]	Semple 7422	1983	Florida	Jefferson Co.	65.4	4578	G03 (2)	2,949,267	0.820
S. tortifolia	Semple 7534	1983	Florida	Brevard Co.	47.6	3332	H03 (2)	1,755,915	0.703
S. tortifolia	Semple and Godfrey 3175	1977	Florida	Holmes Co.	53	3710	A04 (2)	733,905	0.814
S. tortifolia	Kral 41722	1970	Alabama	Geneva Co.	58.6	4102	B04 (3)	746,818	0.825
S. tortifolia	Cook et al., C-669	2001	South Carolina	Berkeley Co.	61.9	4333	C04 (1)	3,236,500	0.657
S. tortifolia[e]	Semple 11833	2010	Georgia	Brooks Co.	29.8	2086	D04 (1)	7,580,752	0.788
S. lepida DC. var. *salebrosa* (Piper) Semple	Semple and Brouillet 4381	1979	Idaho	Boundary Co.	56.3	3941	E04 (3)	1,302,978	0.737
S. lepida var. *salebrosa*	Semple et al., 9209	1990	Wyoming	Carbon Co.	95	6650	F04 (2)	1,542,691	0.709
S. lepida var. *salebrosa*	Semple and Heard 7755	1985	Colorado	Gunnison Co.	47.5	3325	G04 (3)	2,686,967	0.633

Appendix 1. Continued.

Species	Voucher specimen accession no.[a]	Collection year	Collection locality[b]	County[b]	DNA concentration	DNA yield	Gel image well (score)	Quality reads[c]	Missingness[d]
S. lepida var. *salebrosa*	Semple 11154	2003	NW Territories	Nahanni N.P.R.	43.2	2592	H04 (1)	1,243,052	0.710
S. brendiae Semple	Semple and Semple 11432	2006	Quebec	Gaspésie Co.	15	1050	B01 (1)	3,085,092	0.587
S. brendiae	Semple and Semple 11436	2006	Quebec	Gaspésie Co.	29.5	2065	A05 (1)	13,642,478	0.499
S. chilensis Meyen	Lopez Laphitz and Becker 27	2007	Argentina	Catamarca Province	25.6	1792	B05 (1)	1,610,391	0.782
S. chilensis	Lopez Laphitz and Becker 12	2007	Argentina	Chubut Province	34.8	2436	C05 (1)	8,945,242	0.671
S. chilensis	Lopez Laphitz and Becker 10	2007	Chile	Region XI	16.6	1162	D05 (1)	4,456,376	0.709
S. microglossa DC.	Lopez Laphitz and Becker 16	2007	Argentina	Chaco Province	22.4	1120	E05 (1)	5,802,049	0.708
S. microglossa	Lopez Laphitz and Becker 42	2007	Argentina	Chaco Province	66.1	3966	F05 (1)	1,572,344	0.783
S. microglossa	Lopez Laphitz and Becker 41	2007	Argentina	Formosa Province	81.6	5712	G05 (1)	4,912,612	0.712
S. squarrosa Muhl.	Semple 2426	1976	Ontario	Renfrew Co.	37.1	2597	H05 (3)	323,961	0.918
S. squarrosa	Semple 3692	1978	Ontario	Durham Co.	35	2450	A06 (3)	980,080	0.844
S. squarrosa	Cook & Seiden C-125	2000	Quebec	La Vallée-de-la-Gatineau Reg. Co. Mun.	27.6	1932	B06 (2)	2,577,756	0.736
S. bicolor L.	Semple & Chmielewski 5927	1981	Virgina	Nelson Co.	38.1	2667	C06 (3)	491,750	0.885
S. bicolor	Semple & Suripto 9487	1991	Pennsylvania	Perry Co.	63.2	4424	D06 (2)	1,848,949	0.734
S. bicolor	Semple & Brouillet 3614	1978	Connecticut	Hartford Co.	27.2	1904	E06 (2)	1,015,672	0.784
S. bicolor	Semple 4708	1980	New Brunswick	Kent Co.	35.4	2478	F06 (3)	414,189	0.881
S. bicolor	Semple & B. Semple 11472	2006	Prince Edward Island	Queens Co.	23.6	1652	G06 (1)	2,485,719	0.722
S. hispida Muhl. ex Willd. var. *hispida*	Semple & Brouillet 3638	1978	New York	Greene Co.	18.9	1323	H06 (3)	354,252	0.890
S. hispida var. *hispida*	Semple & Keir 4634	1980	Maine	Somerset Co.	37.8	2646	A07 (3)	929,689	0.831
S. hispida × *S. puberula*[e]	Semple, Brammall & Hart 2989	1977	Kentucky	Whitley Co.	26	1820	B07 (3)	196,027	0.924
S. hispida var. *hispida*	Semple & B. Semple 11065	2001	Ontario	Renfrew Co.	32.3	2261	C07 (2)	2,254,614	0.716
S. hispida Muhl. ex Willd. var. *arnoglossa* Fernald	Morton NA12474	1978	Newfoundland	Division No. 5	54	3780	D07 (3)	986,431	0.816
S. hispida var. *hispida*	Semple & Chmielewski 8298	1985	Arkansas	Searcy Co.	35	2450	E07 (3)	593,021	0.866
S. erecta Banks ex Pursh	Semple & Chmielewski 5984	1981	Virginia	Northumberland Co.	21.1	1477	F07 (3)	508,700	0.855
S. erecta	Semple & B. Semple 11189	2003	Tennessee	Coffee Co.	38.3	2681	G07 (2)	540,748	0.843
S. erecta	Semple & Suripto 9501	1991	New Jersey	Atlantic Co.	21.5	1505	H07 (2)	865,169	0.801

Appendix 1. Continued.

Species	Voucher specimen accession no.[a]	Collection year	Collection locality[b]	County[b]	DNA concentration	DNA yield	Gel image well (score)	Quality reads[c]	Missingness[d]
S. erecta	Semple & Suripto 9454	1990	Kentucky	Estill Co.	81.9	5733	A08 (3)	864,935	0.802
S. erecta	Semple & Chmielewski 6098	1981	South Carolina	Chester Co.	29.5	2065	B08 (2)	673,107	0.821
S. erecta	Semple & Suripto 10175	1991	Mississippi	Itawamba Co.	30.7	2149	C08 (2)	303,375	0.893
S. pulverulenta Nutt.	Semple 11635	2006	North Carolina	Pender Co.	38.1	2286	D08 (1)	1,048,859	0.811
S. pulverulenta	Kral 44276	1971	Alabama	Escambia Co.	45.9	3213	E08 (3)	239,754	0.936
S. pulverulenta[e]	Semple & Suripto 10137	1991	Florida	Washington Co.	29.4	2058	F08 (2)	126,047	0.955
S. pulverulenta	Semple & Suripto 9813	1991	South Carolina	Barnwell Co.	39.5	2765	G08 (2)	316,166	0.924
S. puberula Nutt.	Cook & Seiden C-118	2000	Quebec	Vallée-de-l'Or Reg. Co. Mun.	64	4480	H08 (2)	675,139	0.841
S. puberula	Semple & Ringius 7628	1984	Maryland	Kent Co.	43.1	3017	A09 (3)	2,287,448	0.804
S. puberula	Semple 6867	1982	Massachusetts	Worchester Co.	94.4	6608	B09 (3)	648,093	0.859
S. puberula	Semple 10815	1999	North Carolina	Mitchell Co.	69.5	4865	C09 (2)	1,101,645	0.807
S. pallida (Porter) Rydb.	Semple 11304	2004	South Dakota	Pennington Co.	47.5	3325	D09 (2)	1,005,834	0.775
S. pallida	Semple 11401	2006	Wyoming	Crook Co.	33.5	2345	E09 (1)	13,577,446	0.586
S. pallida	Semple & Heard 8082	1985	New Mexico	San Miguel Co.	44.6	3122	F09 (3)	639,439	0.826
S. rigidiuscula (Torr. & A. Gray) Porter	Semple & Zhang 10602	1997	Ontario	Wapole Island	28.3	1981	G09 (2)	8,082,824	0.578
S. rigidiuscula	Semple & Brouillet 4532	1979	Indiana	Porter Co.	24.9	1743	H09 (2)	1,807,560	0.706
S. rigidiuscula	Semple & Chmielewski 9121	1986	Tennessee	Marshall Co.	35.2	2464	A10 (3)	1,447,922	0.766
S. rigidiuscula	Semple & Chmielewski 5063	1980	Wisconsin	Jackson Co.	49.5	3465	B10 (3)	1,166,027	0.770
S. speciosa Nutt.	Semple & Chmielewski 6180	1981	South Carolina	Greenville Co.	27.3	1911	C10 (3)	389,997	0.880
S. speciosa	Semple 11613	2006	Virginia	Mecklenburg Co.	29.3	2051	D10 (1)	956,852	0.789
S. gattingeri Chapm. ex A. Gray	Semple & Chmielewski 5288	1980	Missouri	Camden Co.	80.2	5614	E10 (2)	2,735,608	0.733
S. gattingeri	Dietrich & Jenkins 49	1994	Missouri	Camden Co.	67	4690	F10 (2)	2,790,439	0.732
S. gattingeri	McNeilus 93-1443	1993	Tennessee	Wilson Co.	34.4	2408	G10 (1)	1,413,356	0.732
S. gattingeri	Nordman s.n.	2000	Tennessee	Rutherford Co.	25.2	1764	H10 (1)	1,062,691	0.757
S. gattingeri	Baily s.n.	2000	Tennessee	Rutherford Co.	37.6	2632	A11 (2)	3,257,026	0.679
S. missouriensis Nutt.	Semple & Heard 7699	1985	Colorado	Yuma Co.	68.8	4816	B11 (3)	1,110,380	0.768
S. missouriensis	Semple, Suripto & Ahmed 9195	1990	Nebraska	Lincoln Co.	77.5	5425	C11 (2)	1,566,591	0.732
S. missouriensis	Semple, Suripto & Ahmed 9263	1990	Utah	Cache Co.	144	10080	D11 (2)	2,083,973	0.702
S. missouriensis	Semple & Jeff Semple 8844	1987	Wisconsin	Adams Co.	38.7	2709	E11 (2)	2,286,568	0.692
S. missouriensis	Semple, Suripto & Ahmed 9381	1990	New Mexico	Cibola Co.	155	10850	F11 (2)	1,071,379	0.761
S. missouriensis	Semple & Brammall 2669	1977	Manitoba	Division No. 1	50.5	3535	G11 (2)	2,421,026	0.659

Appendix 1. Continued.

Species	Voucher specimen accession no.[a]	Collection year	Collection locality[b]	County[b]	DNA concentration	DNA yield	Gel image well (score)	Quality reads[c]	Missingness[d]
S. pinetorum Small	Semple & B. Semple 11223	2003	North Carolina	Moore Co.	30.9	1854	H11 (1)	2,265,648	0.668
S. pinetorum	Semple 11625	2006	North Carolina	Hertford Co.	35	1750	A12 (1)	7,173,043	0.592
S. pinetorum	Semple 11599	2006	North Carolina	Rowan Co.	39.6	2772	B12 (1)	6,001,966	0.615
S. pinetorum	Semple & Suripto 9734	1991	North Carolina	Franklin Co.	22.2	1554	C12 (2)	1,021,413	0.771
S. juncea Aiton	Semple 10677	1999	Pennsylvania	Green Co.	32.2	1932	D12 (1)	540,748	0.695
S. juncea	Semple & Keir 4897	1980	Nova Scotia	Hants Co.	26.3	1841	E12 (2)	699,283	0.818
S. juncea	Semple & Brammall 2757	1977	Missouri	Madison Co.	58	2900	F12 (2)	478,402	0.858
S. juncea	Semple & Brammall 2759	1977	Michigan	Berrien Co.	32.5	2275	G12 (2)	890,474	0.789

[a] Vouchers archived at the University of Waterloo Herbarium (WAT), now housed as a unit of the Université de Montréal Herbarium (MT).
[b] State/province; county/administrative unit.
[c] Number of reads containing a full barcode, cut site remnant, and insert sequence.
[d] Percentage of the 8470 unfiltered SNPs missing in the sample.
[e] Samples not analyzed (see text).