Journal of Plant Genetics and Breeding

Devris, J Plant Genet Breed 2023, 7:1

Expert Review Open Access

Welsh Native Conifers and Flowering Plants: DNA Barcoding

Natasha Devris*

National Botanic Garden of Wales, Llanarthne, United Kingdom

Abstract

We present the first national DNA barcode resource for Wales that includes native conifers and blooming plants (1143 species). We have assembled 97.7% coverage for rbcL, 90.2% coverage for matK, and a dual-locus barcode for 89.7% of the native Welsh flora using the plant DNA barcode markers rbcL and matK. For each species, we sampled several individuals, yielding 3304 rbcL and 2419 matK sequences. 85% of our samples come from DNA that was taken from herbarium specimens. Herbarium specimens exhibit lower DNA barcode recoverability than recently obtained material, mostly because of lower amplification success. However, this is offset by the greater sampling effectiveness of species that have already been gathered, identified, and certified by taxonomists. Four methods are used to evaluate the usefulness of DNA barcodes for identification (degree of discrimination), including pairwise and multiple alignments, Neighbour-Joining trees, sequence similarity in BLASTn searches, and the presence of a barcode gap [1]. These methods produce comparable outcomes, with relative discrimination values of 98.6 to 99.8% of genera and 69.4 to 74.9% of all species utilising both markers. Spatially explicit sampling can further enhance species discrimination. Within 1010 km squares, the mean species discrimination using barcode gap analysis (with a multiple alignment) is 81.6%, and within 22 km squares, it is 93.3%. Our library of DNA barcodes for native Welsh conifers and flowering plants gives the most thorough coverage of any national flora and provides a useful framework for a variety of applications that call for precise species identification.

Keywords: DNA Barcoding; Archaeophyte; DNA barcode markers; rbcL; matK

Introduction

Plant species identification is crucial for maintaining and using biodiversity, but it may be hampered by a lack of taxonomic knowledge. Although it may be challenging or impossible to do so using conventional morphological methods, it is occasionally useful to be able to identify species from material such as roots, seeds, pollen, or in mixtures of plants obtained from the air, soil, or water. Although numerous molecular methods have been employed to address this, a global DNA barcoding campaign to standardise molecular identifications using internationally recognised standards and DNA regions has been launched in response to a growing desire for harmonisation and enhanced efficiency [2].

The Plant Working Group (PWG) of the Consortium for the Barcoding of Life (CBOL) recommended that sections of the plastid genes rbcL and matK be adopted as the standard plant DNA barcodes, with the recognition that additional markers may be necessary [3]. This was done after evaluating a number of candidate loci. Using these markers and others, a number of studies have assessed the value of plant DNA barcodes in a taxonomic and floristic context, and a number of applications have been developed that demonstrate the broad potential for plant DNA barcoding [4]. For instance, barcoding techniques have been used to characterise the plant origins of honey and to verify the authenticity of a variety of plant products, including tea, berries, olive oil, tea , and medicinal plants through . Ecological uses have been as varied, ranging from invasive species identification to characterising the diversity of below-ground plants using roots to reconstructing previous vegetation and climate from plant remains in the soil. For use in phylogenetic community ecology, phylogenetic trees have been constructed using sequences obtained through DNA barcoding. The use of DNA barcoding as a tool for identification depends on the development of reliable sequence reference databases. The capacity to use the enormous number of specimens already present in herbaria and access the right taxonomic expertise are crucial for achieving this. A regional floristic context is where DNA barcoding is most beneficial. The creation of floristic DNA barcode databases makes it possible to combine this source of information with other datasets like records of plant distribution and abundance, habitat information, and conservation priorities. Numerous ecological and conservation-based DNA barcoding applications are applicable within a floristic context. Since a geographically constrained sampling typically includes less closely related species than a full sample, a floristic approach may enable greater degrees of species discrimination [5].

The development of a DNA barcode database for all of Wales' native and archaeophyte (species naturalised before 1500 AD) flowering plants, conifers, and trees is described in this article. Wales serves as a perfect case study to show the possibilities of plant DNA barcoding in a floristic setting. It is a small nation (22,000 sq km), having 1143 species of native and archaeophyte seed plants distributed among 455 genera, 95 families, and 34 orders. Since botanical records have been kept for a long time, the country's flora has been thoroughly examined, and the national herbarium has a large collection of the species, including many new additions. For the complete species assemblage, taxonomic knowledge is accessible [6]. The DNA barcode database can be integrated with a variety of national databases that include characteristics like plant distribution and habitat preferences.

Methods and analysis

Sample collection

With aggregate species for the apomictic groupings of Rubus, Taraxacum, and Hieracium, the Welsh flora consists of 1143 native

***Corresponding author:** Natasha devris, National Botanic Garden of Wales, Llanarthne, United Kingdom, E-mail: natashadevris@gardenofwales.org.uk

Received: 02-Jan-2023, Manuscript No. jpgb-22-84922; **Editor assigned:** 05-Jan-2023, PreQC No. jpgb-22-84922; **Reviewed:** 19-Jan-2023, QC No. jpgb-22-84922; **Revised:** 24-Jan-2023, Manuscript No. jpgb-22-84922 (R); **Published:** 31-Jan-2023, DOI: 10.4172/jpgb.1000137

Citation: Devris N (2023) Welsh Native Conifers and Flowering Plants: DNA Barcoding. J Plant Genet Breed 7: 137.

and archaeophyte species of flowering plants and conifers (455 genera, 95 families, and 34 orders). 4272 individuals in total were sampled, including 635 live plants that were collected around Wales and 3637 herbarium specimens (NMW). A taxonomic specialist evaluated all herbarium and fresh material, and herbarium vouchers were created for recently obtained plants. A photograph served as the voucher for vulnerable species for whom the gathering of a herbarium voucher was not acceptable due to conservation concerns [7]. Before collecting specimens of newly gathered plants, all essential permits were secured. Before collecting any material from areas designated for nature conservation in Wales, as well as from the land manager and/or owner for private sites, permission was acquired from the Countryside Council for Wales.

DNA extraction, amplification and sequencing

Freshly gathered plant leaf material was dried in silica gel before the DNA was extracted according to manufacturer's instructions using Qiagen DNeasy kits. For herbarium samples, the technique was changed to increase the success of amplification and sequencing. Before disruption using a TissueLyser II (Qiagen) with 3 mm tungsten carbide beads, 80 l of DDT at 0.75 mg/ml and 20 l of Proteinase K at 1 mg/ml (Sigma) were added to the 400 l of AP1 buffer. Following interruption, incubation was prolonged for a further hour at 65°C in the modified AP1 buffer. 15 minutes were added at the conclusion of the AE buffer incubation step [8].

A reaction mixture containing 10 l of 2 Biomix, 0.4 l of F and R primer (10 M), and 0.8 l of BSA at 1 mg/ml was used to amplify DNA 2 l of template DNA and 6.4 l of H2O (Sigma). [95°C 2 min (95°C 30 sec, 50°C 1 min 30, 72°C 40 sec)45 cycles, 72°C 5 min, 30°C 10 sec] were the cycling conditions for the PCR. On 1% agarose gels, PCR results were observed. Samples with sufficiently bright bands were sent to Macrogen Europe (Amsterdam, Europe) for purification and DNA sequencing in both directions on an ABI3730X using the same PCR primers.

Sequence editing, checking and alignment

Both pairwise and multiple alignments were utilised to compare the results obtained utilising the various alignment techniques. The Needleman-Wunsch technique was used to calculate all possible global pairwise alignments, and uncorrected p-distances were calculated for each pair using only clear-cut sequence differences. The rbcL sequences were aligned for the multiple alignments using MUSCLE v3.7 with the default settings. For matK, nucleotides were converted into amino acids using transAlign v1.2, aligned using Clustal W v2.1 and back-translated, and then manually edited using Geneious Pro v5.4.4. . The final lengths of the various alignments were 542 bp for rbcL and 897 bp (with gaps) for matK after the matrices' ends were squared [9]. For each pairwise comparison, the multiple alignment yielded uncorrected p-distances. The HPC Wales supercomputer cluster was used to do all alignments. The rbcL and matK alignments were combined into a single matrix using a Python script, with matK coming last.

Interspecific and intraspecific divergence

For all genera with numerous species per genus, interspecific divergences for rbcL and matK were estimated using uncorrected p-distances (from the multiple alignments). The mean uncorrected p-distance for each individual sampled for that species was employed as one of three measures of intraspecific divergence, and intraspecific "all." To remove biases brought on by uneven sampling across various species, intraspecific "theta" uses only the distances between distinct haplotypes within a species [10]. The maximum intraspecific distance between individuals for that species is known as coalescent depth.

Discrimination

Discrimination was evaluated for 808 species for which rbcL and matK were both sequenced for numerous individuals per species in order to examine the relative discrimination of rbcL, matK, and combination. In order to provide sources of discrimination failure, species with single sequences were included in the studies. This was evaluated on a broad scale as well as for each class of conifers and flowering plants. To test for discrimination between various species, four methods were used. Detection of a barcode gap utilising the CBOL Plant Working Group's techniques, but with multiple and pairwise alignments [11]. Neighbour Joining trees with the Kimura-2-Parameter and 1000 bootstrap replicates were used to form monophyletic groupings, with the number of monophyletic species being recorded at the level of any, >50% and >70% bootstrap support being recorded for all species having multiple individuals. Each sequence is used as both a database and a query in BLASTn searches [12].

We looked at connections between the number of species we had DNA barcoded in each order and the percentage discrimination achieved using the methods outlined above to determine whether the level of discrimination at the order level linked to the number of species within that taxonomic group.

Discrimination at different spatial scales

For each 10 by 10 km square of Wales, as well as the vice-counties of Cardiganshire, Caernarvonshire, and Anglesey at the 2 by 2 km square level, botanical society of the British islands records were obtained. This used 891,756 plant records in total. A Python script was used to build a species list for each square and determine whether there were any barcode gaps (using multiple alignments) between the species. Heatmap.2 from gplots in R was used to visualise levels of discriminating [13].

Discussion

Herbaria are important repositories for preserved plant matter, and together, their collections offer a unique record of the flora of the planet. Therefore, it is not surprising that herbarium specimens constitute a typical source of DNA for research on the evolution of plants. In spite of this, not much research has been done on the impact of post-mortem DNA damage on data quality in herbarium material. According to our knowledge, this is the first work to use an experimental methodology to assess the common damage types in herbarium DNA.

Degree of DNA fragmentation caused by herbarium DNA preservation

Our findings demonstrate that double-stranded breaks are the most observable type of post-mortem damage in herbarium DNA. It's probable that the high temperatures (60–70°C) used to dry herbarium specimens today cause cells to burst quickly, releasing nucleases, ROS, and other cellular enzymes as a result. Similar to necrosis, such physiological conditions often result in the random degradation of DNA into smaller pieces, which appears as a smear on agarose gels. In our experience, oligo-nucleosomal DNA fragmentation, a sign of programmed cell death, is rarely present in DNA extracts from herbarium tissues [14]. (PCD). Known to trigger additional nucleases and ROS that damage DNA, PCD is a highly coordinated death process[Signs of DNA 'laddering' may suggest that the plant material was maintained too slowly and that the plant had endured

protracted (hours, days) periods of abiotic stress during preparation. The best method for preserving plant DNA has been proven to be rapid desiccation of plant matter, as it slows the PCD damaging process.

According to our findings, Taq polymerase may not be able to access up to 89.5% of the DNA in juvenile herbarium specimens. Herbarium extracts were not found to slow down the amplification of an exogenous DNA control, indicating that the poor amplification success was not caused by the presence of PCR inhibitors. Another hypothesis is that there may be significant copy number heterogeneity in herbarium DNA samples and that damage may be locus- or region-specific [15]. However, it was discovered that gene copy numbers inside each genomic compartment were of the same order of magnitude, disproving the theory of region-specific damage. It is true that a two or three locus depiction of each genomic compartment is a crude picture of the intricacy of the organellar and nuclear genomes. To rule out potential hotspots of DNA deterioration in the herbarium, a more representative sampling will be required. Another factor to take into account is the fact that the DNA in the herbarium was extensively altered, either by double stranded breaks, inter-strand cross-links, abasic sites, or other structural changes. Additionally, DNA molecules cannot be amplified or sequenced if they have this kind of polymerase non-bypassable damage. For instance, some research on ancient DNA came to the conclusion that preventing lesions significantly altered DNA molecules. Analyzing the proportion of abasic sites and other polymerase non-bypassable damage in herbarium DNA will be extremely helpful for developing future DNA repair and rescue techniques. These techniques can enhance the retrieval of precise sequence data from herbarium specimens. For instance, certain forms of blocking lesions could be eliminated by pre-treating the herbarium DNA with base excision repair enzymes. Furthermore, the amplification of severely damaged herbarium DNA molecules may be possible with the use of designed polymerases that can stretch through blocking lesions [16].

Degree of DNA sequence modification in herbarium DNA

In comparison to young and fresh tissues, the plastid DNA of old herbarium specimens showed higher levels of C-T/G-A transitions. Therefore, the observed nucleotide substitutions (which we would expect to occur without bias) cannot all be accounted for by polymerase and sequencing errors alone [17]. Instead, additional misincorporations resulting from miscoding lesions present in old herbarium DNA molecules may be a more plausible explanation. Despite this, we discovered that the observed levels were low (21.4% of miscoding lesions), and that the C-T-G-A rate was 1.53 x 106 per nucleotide per year of herbarium storage. The most frequent type of miscoding lesions in ancient DNA are CT/GA transitions, which are caused by the hydrolytic deamination of cytosine to uracil [18]. According to our findings, herbarium specimens are shielded from the primary dangers to their long-term preservation, but their DNA is still vulnerable to hydrolytic activity. It is likely that some rehydration of DNA due to airborne moisture may result in hydrolytic DNA damage.

Conclusion

Our research demonstrates the enormous wealth of trustworthy DNA sequence information available in herbaria. Even though it is more challenging, it is now possible to generate DNA sequence data from badly fragmented DNA using a variety of next-generation sequencing techniques. However, our finding that polymerase-accessible DNA may be reduced by up to 90% may make it difficult to prepare samples using existing NGS techniques, particularly when there is a shortage of herbarium material. The level of sequence alteration would most likely be a more serious problem. Our findings suggest that CT/GA substitutions might possibly result in the production of an inaccurate DNA sequence; however, we forecast the sequence error rate to be insignificant (0.03%), even if PCR products (i.e., the 750-bp rbcL barcode region sequence) The significance of these issues, however, is dependent on the type of inquiry being conducted; for instance, random errors are more likely to cause random noise than a phylogenetic signal. In fact, there is currently no evidence that older herbarium accessions share more type 2 transitions when compared to more recent ones in comparative analyses that include both types of accessions.

Acknowledgement

None

Conflict of Interest

None

References

1. Bebber DP, Carine MA, Wood JRI, Wortley AH, Harris DJ, et al. (2010) Herbaria are a major frontier for species discovery. Proc Natl Acad Sci USA 107: 22169-22171.
2. Sebastian P, Schaefer H, Telford IRH, Renner SS (2010) Cucumber (Cucumis sativus) and melon (C. melo) have numerous wild relatives in Asia and Australia, and the sister species of melon is from Australia. Proc Natl Acad Sci USA 107: 14269-14273.
3. Metzker ML (2010) Sequencing technologies – the next generation. Nat Rev Genet 11: 31-46.
4. Cubero OF, Crespo A, Fatehi J, Bridge PD (1998) DNA extraction and PCR amplification method suitable for fresh, herbarium-stored, lichenized, and other fungi. Pl Sys Evol 216: 243-249.
5. Drabkova L, Kirschner J, Vlcek C (2002) Comparison of seven DNA extraction and amplification protocols in historic herbarium specimens of Juncaceae. Pl Mol Biol Rep 20: 161-175.
6. Telle S, Thines M (2008) Amplification of cox2 (□620 bp) from 2 mg of up to 129 years old herbarium specimens, comparing 19 extraction methods and 15 polymerases. Plos One 3: 3584.
7. Doyle JJ, Dickson EE (1987) Preservation of plant species for DNA restriction endonuclease analysis. Taxon 36: 715-722.
8. Pyle MM, Adams RP (1989) In situ preservation of DNA in plant specimens. Taxon 38: 576-581.
9. Harris SA (1993) DNA analysis of tropical plant species: an assessment of different drying methods. Pl Syst Evol 188: 57-64.
10. Chase MW, Hills HH (1991) Silica gel: an ideal material for field preservation of leaf samples for DNA studies. Taxon 40: 215-220.
11. Srinivansan M, Sedmak D, Jewell S (2002) Effect of fixatives and tissue processing on the content and integrity of nucleic acids. Am J Pathol 161: 1961-1971.
12. Roldán-Arjona T, Ariza RR (2009) Repair and tolerance of oxidative DNA damage in plants. Mut Res 681: 169-179.
13. Boesch P, Weber-Lotfi F, Ibrahim N, Tarasenko V, Cosset A, et al. (2011) DNA repair in organelles: pathways, organization, regulation, relevance in disease and aging. Biochim Bioph Acta 1813: 186-200.
14. Stiller M, Green RE, Ronan M, Simons JF, Du L, et al. (2006) Patterns of nucleotide misincorporations during enzymatic amplification and direct large-scale sequencing of ancient DNA. Proc Natl Acad Sci USA 103: 13578-13584.
15. Hollingsworth PM, Forrest LL, Spouge JL, Hajibabaei M, Ratnasingham S, et al. (2009) A DNA barcode for land plants. Proc Natl Acad Sci USA 106: 12794-12797.
16. Proost S, Van Bel M, Sterck L, Billiau K, Van Parys T, et al. (2009) PLAZA: a comparative genomics resource to study gene and genome evolution in plants. Pl Cell 21: 3718-3731.

17. Goecks J, Nekrutenko A, Taylor J, The Galaxy Team (2010) Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. Genome Biol 25: 86.

18. Hansen AJ, Willerslev E, Wiuf C, Mourier T, Arctander P (2001) Statistical evidence for miscoding lesions in ancient DNA templates. Mol Biol Evol 18: 262-265.