Metadata for:

**Towards an integrative strategy for conservation of crop wild relatives (CWR) and wild-utilized plants (WUS) in Canada**

Jens C. Ulrich\*, Tara L. Moreau, Erika Luna-Perez, Kephra I.S. Beckett, Lili K. Simon, Zoë Migicovsky, Axel Diederichsen, and Colin K. Khoury

Prepared February 28, 2022 by Jens Ulrich

# Supplemental Table S1 – Inventory

Format - .csv file

TAXON – name of the taxon, either a species or infraspecific taxon. For species that include subordinate infraspecific taxa, rows are listed for each taxonomic level. For example, the row TAXON == “Acer glabrum” is immediately followed by the row with its subordinate TAXON == “Acer glabrum var. douglasii”. Full scientific name (i.e., name including authorship) is reported in the column scientificName (e.g., scientificName == “Acer glabrum var. douglasii (Hook.) Dippel”). To remove duplicate taxa, the inventory must be sorted by FINEST\_TAXON\_RESOLUTION == “Y”. Sorting would remove e.g., the row where TAXON == “Acer glabrum” but not e.g, the row where TAXON == “Acer saccharum” given that Acer saccharum does not include any subordinate infraspecific levels. Such filtering **must** be done to reproduce counts of taxa in the paper. I.e., the .csv file includes 935 rows but only 779 taxa, once double-dipped taxa such as Acer glabrum and Acer glabrum var. douglasii are filtered to a single row.

PRIMARY\_ASSOCIATED\_CROP\_COMMON\_NAME – Common name of associated crop, e.g. *Chenopodium berlandieri* is related to the crop “Quinoa”. OR for non-crop relatives (i.e., wild-utilized plants that are not also recognized as CWR), e.g., *Mahonia spp.* are collectively commonly referred to as “Barberry”.

SECONDARY\_ASSOCIATED\_CROP\_COMMON\_NAME – In select cases, we list the name of a secondary crop relationship. Canada’s CWR diversity includes several CWR from the genus *Helianthus* which are relatives of both Sunflower and Jerusalem-artichoke. For the species that are closely related to Jerusalem-artichoke (and the relationship status given in the column GENEPOOL refers to this relationship) we provide the name of this more closely allied species. Similar for CWR from the genus *Vaccinium* that are more closely related to cranberry versus the various blueberry crop species, we list the associated crop name of cranberry here to convey that the close genepool status is specific to its relationship to cranberry and not to highbush/lowbush/rabbits eye blueberry.

GENUS – genus name

SPECIES – species name; infraspecific taxa from the same species are united through shared values in this column.

RANK – if row is for an infraspecific taxon, rank specifies the designation for the unit, either var. (variety) or subsp. (subspecies), e.g. for “Acer glabrum var. douglasii”, RANK == “var.”

INFRASPECIFIC – the infraspecific epithet, e.g. for “Acer glabrum var. douglasii”, INFRASPECIFIC == “douglasii”

scientificName – full scientific name of the taxon including authorship

INFRASPECIFIC\_COLLECTIONS\_ASSIGNED – Generally, infraspecific identifications for accessions are either not assigned OR we are not confident that they are assigned with accuracy by expert taxonomist. If “Y” we had manually assigned infraspecific taxon names to all accessions within the species (collected from within Canada). We only did so if either a) only 1 infraspecific taxon exists within Canada and thus couldn’t be confused or b) if infraspecific taxa are geographically isolated and therefore there is little chance of mistaking the identifications. If “N” no infraspecific taxon names are given to accessions within the species, all accessions are labelled at the species only. If blank/NA the taxon is a species with no subordinate taxa.

FINEST\_TAXON\_RESOLUTION – Use this column as a way to filter out double-dipped taxa – e.g. don’t want to count both Acer glabrum and Acer glabrum var. douglasii if a count of the number of relatives of sugar maple is desired. In this case we would only want these two taxon levels to count as a single CWR of sugar maple.

CWR – If “Y” taxon is a crop wild relative (CWR). We only scored as crop wild relatives if the species is drawn from a genus with a crop listed in the GRIN – crop wild relative database. We expanded the definition to include 3 genera as food crops due to national importance in Canada that are not included in the GRIN list: saskatoon, flax, and sugar maple. All species in these genera are therefore also scored as CWR. We also listed as CWR any taxa that are cultivated forest resources, ornamentals, or forage and feed crops (or in a genus with a species that is one of these). Species that are wild-utilized but not cultivated on any economic scale are not scored as CWR.

WUS – if “Y” taxon is a wild-utilized species. Not mutually exclusive with the CWR designation.

NATIVE – native or naturalized with established populations in Canada, according to the Canadensys database of vascular plants of Canada <https://community.canadensys.net/>

PRIMARY\_ASSOCIATED\_CROP\_TYPE\_GENERAL\_1 – The category of CWR or WUS use.

PRIMARY\_ASSOCIATED\_CROP\_TYPE\_GENERAL\_2 – Alternate category of CWR or WUS use.

PRIMARY\_CROP\_OR\_WUS\_USE\_SPECIFIC\_1 – The subcategory if a food crop relative OR if not a food crop relative == PRIMARY\_ASSOCIATED\_CROP\_TYPE\_GENERAL\_1.

PRIMARY\_CROP\_OR\_WUS\_USE\_SPECIFIC\_2 – Alternate specific use. E.g., some plants such as *Crataegus spp.* produce edible berry fruits but are also forest genetic resources.

PRIMARY\_CROP\_OR\_WUS\_USE\_SPECIFIC\_3 - Alternate specific use. E.g., some plants such as *Crataegus spp.* produce edible berry fruits but are also forest genetic resources.

TIER – if a food crop relative == 1; if other type of crop wild relative (not directly related to food crop and food crop security) == 2; if a WUS that is not also a CWR, == NA

GENEPOOL – genetic distance to a crop species, following the classification scheme of Harlan & de Wet 1971. A = primary relative/progenitor/crop species, B = secondary relative, C = distant relative or unknown genetic distance.

CATEGORY – Prioritization based on breeding utility. == TIER + GENEPOOL **except** when taxon has value as graftstock, the GENEPOOL score is overridden and the prioritization category is assigned to highest value, 1A. Tier 2 CWR do not follow the same genepool scheme (generally there is no single target crop, e.g. for the genus Pinus – most species are utilized as forest resources) and are all assigned category == 2. WUS are not differentiated by category.

GRAFTSTOCK – known value as graftstock for cultivated crops.

TRAITS – multiple columns allowing for single trait per column (obtained via GRIN – CWR database).

ROUNDED\_G\_RANK – global conservation score from low conservation concern G5 to highest conservation concern G1. Obtained from NatureServe

ROUNDED\_G\_RANK – national (Canadian) conservation score from low conservation concern N5 to highest conservation concern N1. Obtained from NatureServe

CURRENT\_US\_DISTRIBUTION – states where the taxon is known to occur.

CURRENT\_CA\_DISTRIBUTION – provinces where the taxon is known to occur.

COSEWIC\_DESC – cosewic assessment score (if taxon has been assessed by the COSEWIC panel).

SOURCE\_OF\_TAXON – source where the taxon was obtained during the inventory assessment process (see Methods and Materials).

WUS\_SOURCE – source used to score whether a taxon is wild-utilized.

# Supplemental\_Table\_S2\_out\_of\_bound\_ranges\_provincial

Format: .csv fie

We used this file to filter the GBIF occurrence points to remove regions for species that are outside of their known ranges but are represented by valid occurrence points. Likely these points represent specimens cultivated in gardens, cities, etc rather than representing established persistent populations.

SPECIES – species with regions recognized by GBIF occurrences that are outside it’s range

PROVINCE – name of the province outside of the range for which occurrence points were removed.

SOURCE – source for identifying that the region does not truly contain persistent natural pops.

# Supplemental\_Table\_S3\_out\_of\_bound\_ranges\_ecoregions

Format: .csv fie

We used this file to filter the GBIF occurrence points to remove regions for species that are outside of their known ranges but are represented by valid occurrence points. Likely these points represent specimens cultivated in gardens, cities, etc rather than representing established persistent populations.

SPECIES – species with regions recognized by GBIF occurrences that are outside it’s range

ECOREGION – name of the ecoregion outside of the range for which occurrence points were removed.

SOURCE – source for identifying that the region does not truly contain persistent natural pops.

# Supplemental\_Table\_S4\_ summary\_accessions\_all\_species

Format: .csv fie

Includes some informational columns from the inventory and ALSO the following 6 columns summarizing the number of accessions in genebanks.

total\_accessions\_sp – the total number of accessions (at the species taxonomic level) regardless of provenance (could be collected from outside Canada, e.g. in the United States).

garden\_accessions\_sp – the number of accessions in botanical gardens (at the species taxonomic level) regardless of provenance (could be collected from outside Canada, e.g. in the United States).

genebank\_accessions\_sp – the number of accessions in genebanks (both Canada’s national ex situ PGR system and the United States NPGS) (at the species taxonomic level) regardless of provenance (could be collected from outside Canada, e.g. in the United States).

total\_accessions\_w\_finest\_taxon\_res – the total number of accessions (at the finest taxonomic resolution possible) including only those that are wild-collected from populations in Canada. We were unable to assign infraspecific taxon id’s for some subspecies taxa (column INFRASPECIFIC\_COLLECTIONS\_ASSIGNED == “N”) – in these cases the numbers for the infraspecific taxa are listed as NA and one must read the higher-level of the species in which the infraspecific taxon is nested. In cases where infraspecific taxa are assigned, the value in this column (and the garden and genebank counts) are sums of all subordinate taxa.

garden\_accessions\_w\_finest\_taxon\_res – the number of accessions in botanical gardens (at the finest taxonomic resolution possible) including only those that are wild-collected from populations in Canada.

genebank\_accessions\_w\_finest\_taxon\_res – the number of accessions in genebanks (both Canada’s national ex situ PGR system and the United States NPGS) (at the finest taxonomic resolution possible) including only those that are wild-collected from populations in Canada.