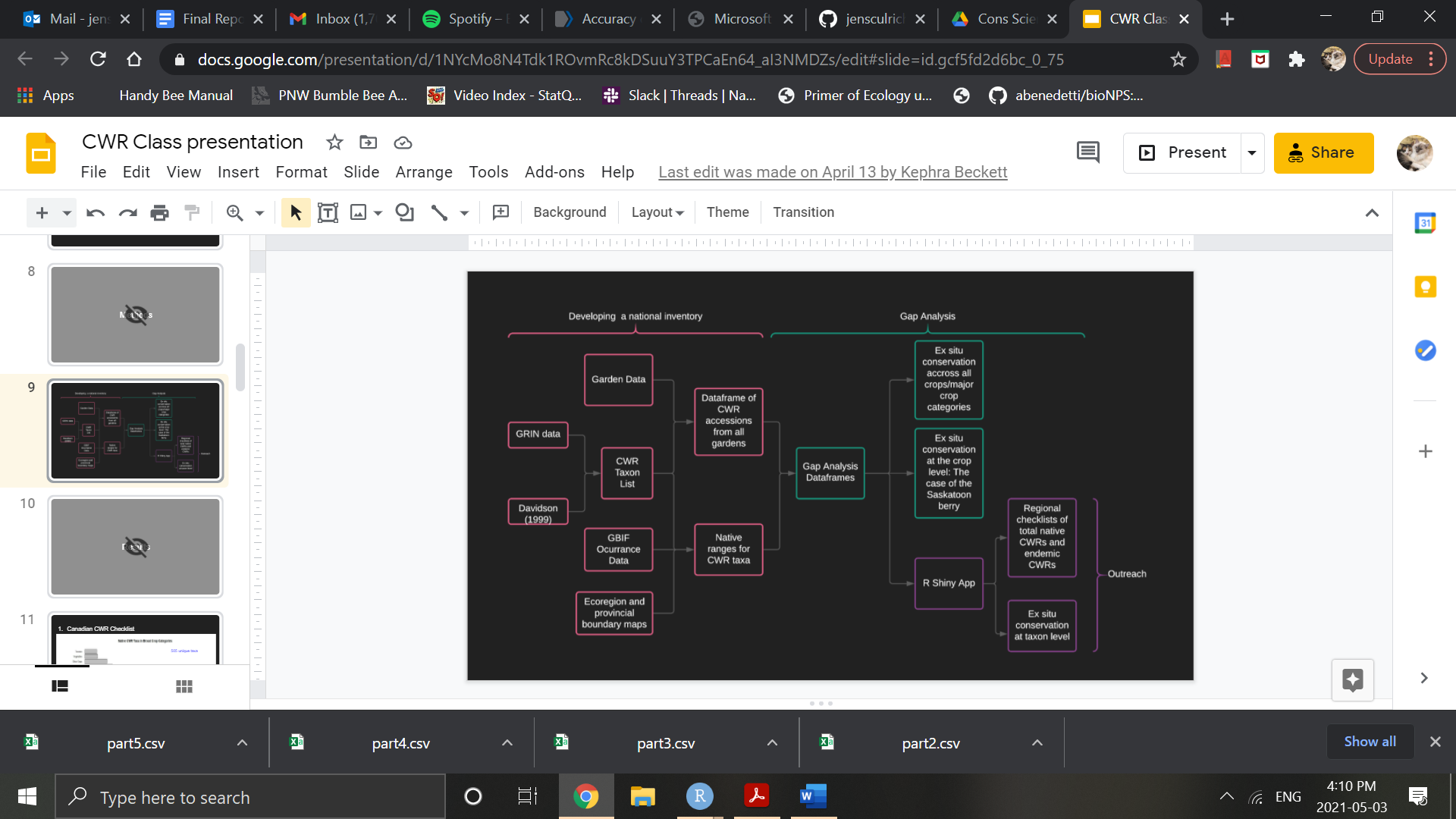
***Metadata – CWR Project***

Updated May 3, 2021

Workflow Outline:



**CWR\_checklists**

This subdirectory includes lists of CWRs from 4 different sources. Ultimately, we combined the lists from Davidson 1995 and from the GRIN database (filtered to plants that occur in Canada) to produce our list of CWR taxa.

Davidson\_Canada\_raw\_data.csv

Davidson, C. G. (1995). Canadian wild plant germplasm of economic significance. *Canadian Journal of Plant Science*, *75*(1), 23–32. https://doi.org/10.4141/cjps95-006

GRIN\_Canada\_cleaned\_data.csv

Downloaded from: https://npgsweb.ars-grin.gov/gringlobal/search

GRIN\_US\_cleaned\_data.csv

Downloaded from: https://npgsweb.ars-grin.gov/gringlobal/search

pnas\_tableS1\_USinventory\_1A\_1B\_1C.csv

Khoury, C. K., Carver, D., Greene, S. L., Williams, K. A., Achicanoy, H. A., Schori, M., León, B., Wiersema, J. H., & Frances, A. (2020). Crop wild relatives of the United States require urgent conservation action. *Proceedings of the National Academy of Sciences*, *117*(52), 33351–33357. https://doi.org/10.1073/pnas.2007029117

**Garden\_Data**

This subdirectory includes garden data from all the gardens that participated in our study, with individual .csv files from each garden. These data have already been filtered down to match those species in our CWR list.

**GBIF\_download\_inputs**

This folder contains 6 separate csv’s containing all of the taxon keys used to match our CWR species with taxon identifiers used by GBIF. Data was kept in 6 separate files since the number of occurrence downloads is capped at 100,000 and will exceed if all of the taxon keys are kept together.

**Geo\_Data**

This subdirectory includes the input geographic data that, after processing, was used to define the region of occurrence for natural occurrences from GBIF and for wild origin garden accessions, and was also used for graphical mapping.

Canada.geojson

Downloaded from: ht<tps://www12.statcan.gc.ca/census-re>censement/2011/geo/bound-limit/bound-limit-2011-eng.cfm

Used this Canada layer to clip ecoregions of the world (World\_ecoregions.geojson) to Canada’s national boundary.

Canada\_ecoregions\_clipped.geojson

World ecoregions file after clipping to Canada’s national border.

Canada\_provinces.geojson

Downloaded from: ht<tps://www12.statcan.gc.ca/census-re>censement/2011/geo/bound-limit/bound-limit-2011-eng.cfm

World\_ecoregions.geojson

Ecoregion level 3, downloaded from: http://maps.tnc.org/gis\_data.html

**Input\_Data\_and\_Files**

garden\_contact\_list\_full.csv

Includes all of the gardens that we attempted to contact, including contact info

garden\_list.csv

Includes only the gardens that contributed full, usable data (for graphical mapping)

master\_list\_apr\_3.csv

Our list of CWR taxa, which includes species identified from multiple resources (see CWR Checklists subdirectory). Crop categories and IUCN conservation status were designated and added by our working group.

**Output\_Data\_and\_Files**

anti\_join\_missing\_cwrs.csv

A list of CWRs that we could not produce range maps for using the GBIF occurrence data.

ecoregion\_gap\_table.csv

This table includes a row for each unique ecoregion in which each unique CWR naturally occurs. When Garden = NA for a row, no surveyed garden had an accession that was wild collected from that native region. When Garden != NA, the named garden identifies the collection where an accession of origin from that region occurs. When there are more than one wild origin accessions from a unique ecoregion for a CWR, the row was expanded. For example, if three accessions from native region X were identified for CWR species Y, the table would include 3 unique rows for that CWR x region combination.

ecoregion\_gap\_table\_post\_manual\_range\_edits.csv

This table is a manually edited version of ecoregion\_gap\_table.csv. As outlined in our methods, we manually removed rows that represent ecoregions outside of the likely native range of each CWR.

province\_gap\_table.csv

This table includes a row for each unique province in which each unique CWR naturally occurs. When Garden = NA for a row, no surveyed garden had an accession that was wild collected from that native region. When Garden != NA, the named garden identifies the collection where an accession of origin from that region occurs. When there are more than one wild origin accessions from a unique ecoregion for a CWR, the row was expanded. For example, if three accessions from native region X were identified for CWR species Y, the table would include 3 unique rows for that CWR x region combination.

province\_gap\_table\_post\_manual\_range\_edits.csv

This table is a manually edited version of province\_gap\_table.csv. As outlined in our methods, we manually removed rows that represent ecoregions outside of the likely native range of each CWR.

saskatoon\_gap\_table\_ecoregion.csv

This table is a filtered version of ecoregion\_gap\_table.csv., with filter(Crop == “Saskatoon”).

saskatoon\_gap\_table\_province.csv

This table is a filtered version of province\_gap\_table.csv., with filter(Crop == “Saskatoon”).

**Rfile\_unite\_Davidson\_and\_GRIN\_CWR\_lists.R**

Use this file to read, clean and fully join different CWR lists into a single list that includes both matching and unique taxa from each input list.

**Rfile\_identify\_cwr\_accessions\_from\_garden\_collection\_data.R**

Use this file to filter and clean garden collection data so that it only includes CWRs (rather than full garden collection) and is formatted for combining with other gardens into a full inventory.

**Rfile\_GBIF\_download\_clean.R**

Use this file to find occurrence points from GBIF that we used to define species ranges. The output download is large and cannot be stored in a standard github directory.

**Rfile\_process\_GBIF\_data\_for\_native\_range\_and\_gap\_analyses.R**

Use this file to combine all garden data into one full inventory, and then generate gap tables by joining the garden inventory data frame with the natural occurrence data.

**Rfile\_conduct\_native\_occurrence\_summary\_statistics.R**

Use this file to identify regional CWR species richness.

**Rfile\_conduct\_gap\_analyses.R**

Use this file to conduct the geographical gap analysis on the gap tables (by running a function to determine proportion of native ranges represented in gardens for each species).

/Input\_Data\_and\_Files/long\_GBIF\_2.json