Exploration of **FELLOW** trait dataset

The objective of this document is to:

- visually explore the trait database
- understand trait coverage / data gaps
- check for possible inconsistencies

Description of the species list

We compiled the species lists from 16 datasets. After cleaning and harmonization, there were 1706 unique taxa.

FAMILY	GENUS	SPECIES	SUBSPECIES	VARIETY
12	197	1430	62	5

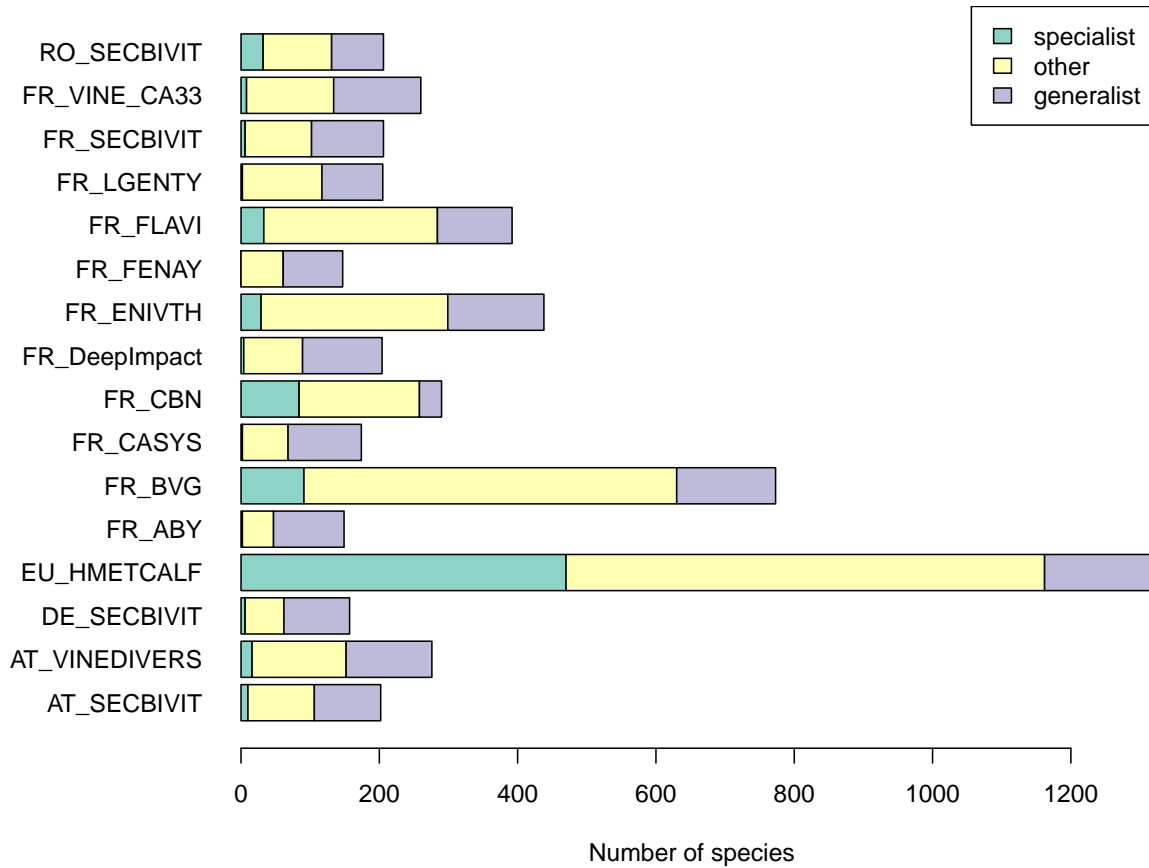Let's define:

specialist: a taxa that occured only in a singe database

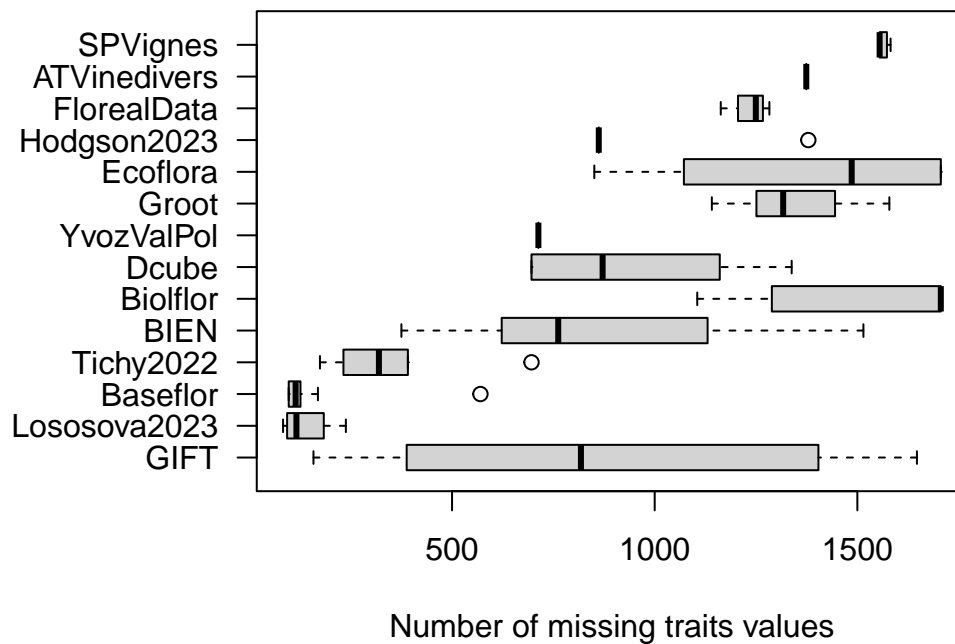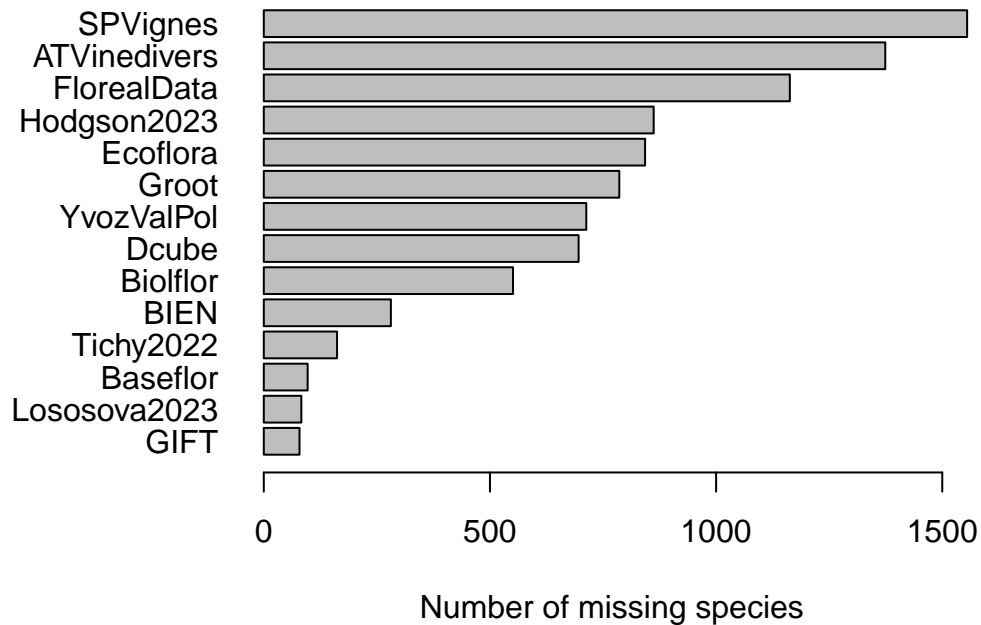*generalist*: a taxa that is listed in 50% of the databases (8 out of 16)

```
sp_class
specialist     other generalist
     773        789         144
```
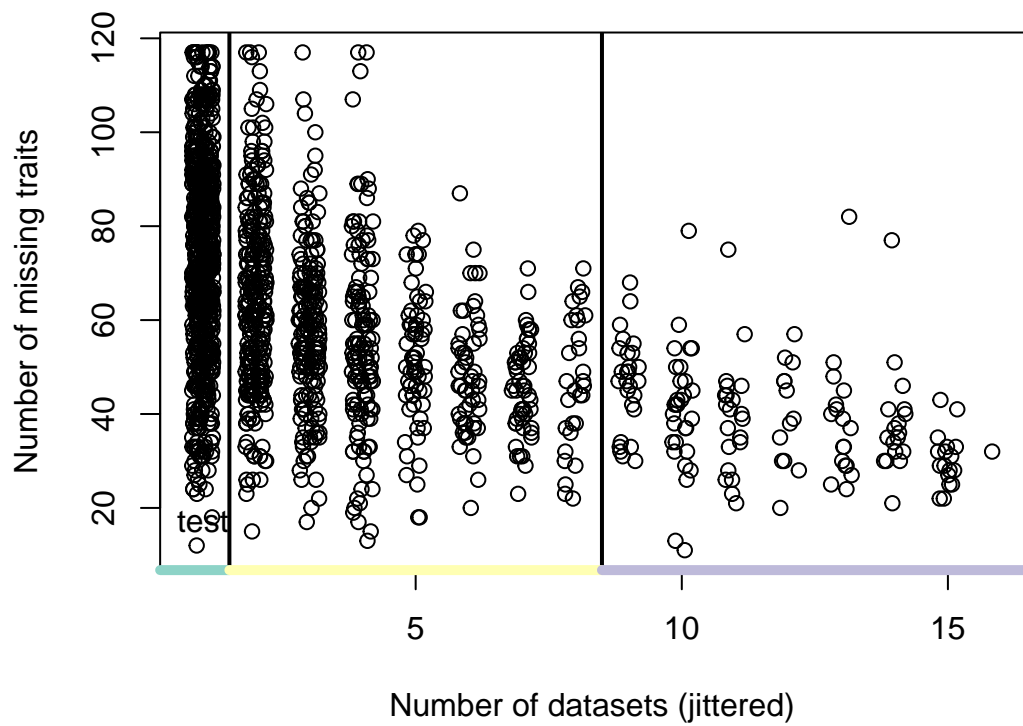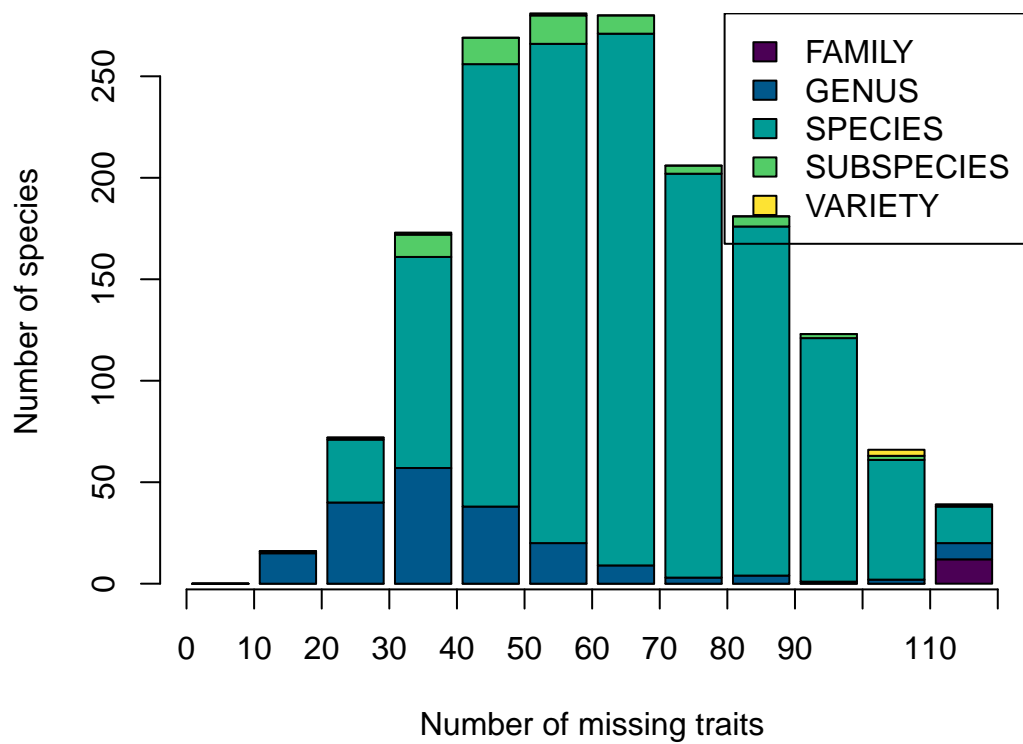

Number of species

Description of trait databases

So far, we compiled 117 traits for 1706 taxa gathered from 14 trait databases. But there are many missing values.

Number of missing species

Number of missing traits values

Taxa with no or limited trait information (N=20).

```
 [1] "Acacia"               "Agrimonia agrimonoides" "Agropyron"
 [4] "Amaranthaceae"        "Apiaceae"               "Asparagaceae"
 [7] "Brassicaceae"         "Caryophyllaceae"        "Crambe abyssinica"
[10] "Dysphania aristata"   "Glyceria"               "Lamiaceae"
[13] "Liliaceae"            "Paronychia"             "Piptatherum"
[16] "Poaceae"              "Roemeria hispida"       "Rosaceae"
[19] "Rubiaceae"            "Viburnum"
```
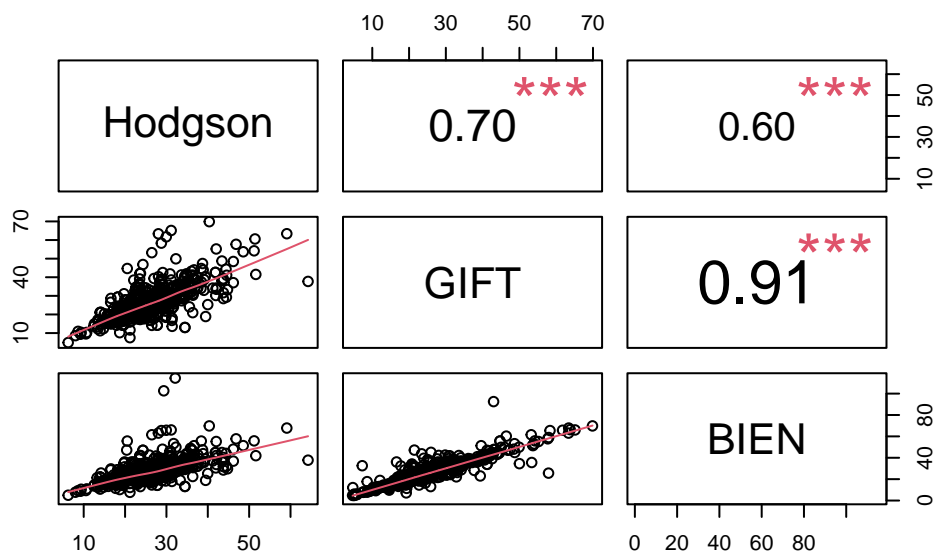
Open question:

How to deal with families taxa?

How to deal with missing trait values? Trait imputation, discarding taxa, …

Comparison

SLA

There are three sources of information for Specific leaf area (SLA) : Hodgson et al. 2023 (in mm2/mg), GIFT (in cm2/g) and BIEN (in m2/kg = mm2/mg).
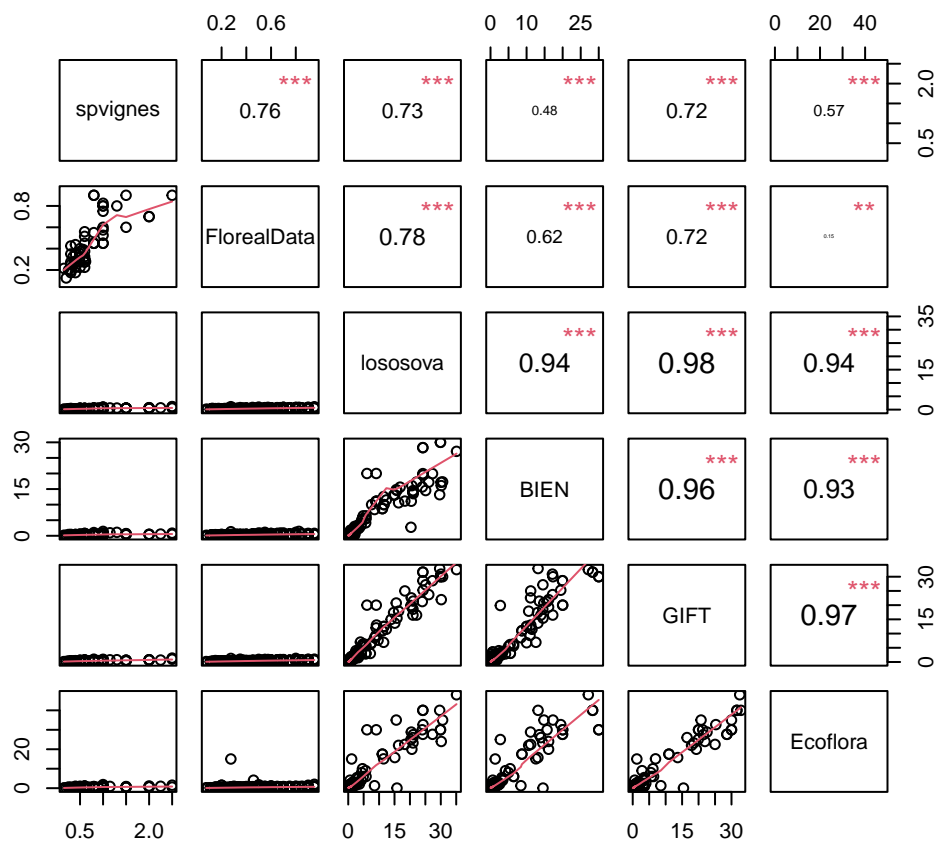

Values are highly correlated, so we could imagine filling the missing values (using preferred data sources or averaging them).

Number of NAs:

```
Hodgson    GIFT    BIEN  filled
     862     731     720     475
```

Plant height

There are six sources of information for plant height.
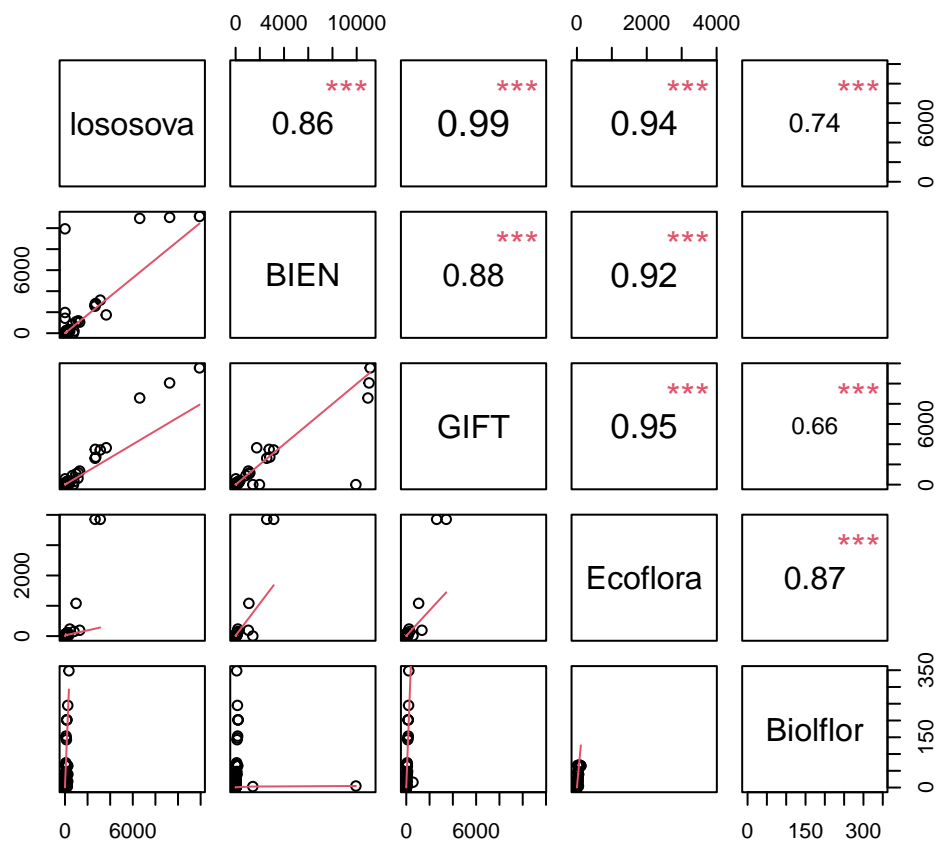

SPVignes and FlorealData are limited to small plants (<1m) (no trees) *but FlorealData>100cm must be clean.*

Number of NAs:

spvignes	FlorealData	lososova	BIEN	GIFT	Ecoflora
1582	1345	129	610	492	851

filled
106

Seed mass

There are five sources of information for seed mass

lososova	BIEN	GIFT	Ecoflora	Biolflor	filled
238	589	284	1072	1289	215

Flower colour

There are five sources of information for flower colour, but it must be cleaned

```
  Baseflor        BIEN      GIFT FloralData   Biolflor
       169        1483      1444        1251       1706
```

```
 [1] "Blanc"
 [2] "Blanc_Blanc, jaune_Bleu"
 [3] "Blanc_Blanc, jaune, bleu_Bleu_Bleu, jaune"
 [4] "Blanc_Blanc, rose"
 [5] "Blanc_Blanc, rose_Bleu_Jaune_Rose"
 [6] "Blanc_Blanc, rose_Bleu_Rose"
 [7] "Blanc_Blanc, rose_Jaune_Rose"
 [8] "Blanc_Blanc, rose_Rose"
 [9] "Blanc_Bleu_Bleu, blanc"
[10] "Blanc_Bleu_Bleu, blanc_Jaune"
[11] "Blanc_Bleu_Bleu, rose"
[12] "Blanc_Bleu_Jaune_Rose"
[13] "Blanc_Jaune"
[14] "Blanc_Jaune_Rose"
[15] "Blanc_Jaune_Vert"
[16] "Blanc_Rose"
[17] "Blanc_Vert"
[18] "Blanc, jaune"
[19] "Blanc, jaune_Bleu, jaune, rose"
[20] "Blanc, jaune_Jaune"
[21] "Blanc, jaune, bleu"
[22] "Blanc, jaune, rose"
[23] "Blanc, rose"
[24] "Blanc, rose_Blanc, vert, rose_Rose"
[25] "Blanc, rose_Jaune_Jaune, rose_Rose"
[26] "Blanc, rose_Rose"
[27] "Blanc, rose_Vert"
[28] "Blanc, vert, rose"
[29] "Bleu"
[30] "Bleu_Bleu, blanc, rose_Bleu, jaune_Bleu, rose"
[31] "Bleu_Bleu, blanc, rose_Bleu, rose_Jaune_Rose"
[32] "Bleu_Bleu, rose"
[33] "Bleu_Jaune"
[34] "Bleu_Jaune_Rose"
[35] "Bleu_Rose"
```

```
[36] "Bleu_Vert"
[37] "Bleu, blanc"
[38] "Bleu, blanc, rose"
[39] "Bleu, jaune"
[40] "Bleu, jaune, rose"
[41] "Bleu, rose"
[42] "Jaune"
[43] "Jaune_Marron_Noir_Vert"
[44] "Jaune_Rose"
[45] "Jaune_Vert"
[46] "Jaune, rose"
[47] "Jaune, vert"
[48] "Marron"
[49] "Marron_Rose_Vert"
[50] "Marron_Vert"
[51] "Noir"
[52] "Rose"
[53] "Rose_Vert"
[54] "Rose_Vert_Vert, rose"
[55] "Rose_Vert, jaune, rose"
[56] "Vert"
[57] "Vert_Vert, bleu"
[58] "Vert_Vert, bleu_Vert, rose"
[59] "Vert_Vert, rose"
[60] "Vert, bleu"
[61] "Vert, jaune, rose"
[62] "Vert, rose"
```