

- Die Zeilen 1 und 2 sind Info-Zeilen was in der Datei abgelegt wird
- Zeile 3 (und weitere Zeilen mit # beginnend) markieren den Anfang eines neuen paarweisen Vergleichs zwischen Genen zweier Spezies.
- Zeile 4,5 sind jeweils eine 'Kante' mit reciprocal best matches; im Detail:
 - Spalte 1 und 2: Namen der Gene aus den Eingabedateien: E_10 kommt aus E.faa und C_10 kommt aus C.faa. *Achtung*: es handelt sich bei dem Gennamen um E_10 in der Datei E.faa.
 - Spalte 3,4: Evalue und Bitscore in Richtung a -> b
 - Spalte 5,6: Evalue und Bitscore in Richtung b -> a

```
# file_a    file_b
# a b    evalue_ab    bitscore_ab evalue_ba    bitscore_ba
# E.faa C.faa
E_10    C_10    1.21e-143    425 4.59e-136    405
E_11    C_11    1.50e-59    185 1.66e-59    185
```

Hinweise

- Gennamen muessen nicht eindeutig sein
- Beim Parsen jeweils eines neuen paarweisen Vergleiches (# E.faa C.faa) soll deshalb der Dateiname mitgenommen werden und jede Kante soll so codiert sein: (E.faa,E_10) -> (C.faa, C_10). Die genaue Codierung ist euch ueberlassen (Tupel, Strukturen, etc)
- Darauf achten das sowohl bitscore, evalue, als auch similarity (bei den ffdj Dateien) gute Kantengewichtungen sind
- evalue ist besser wenn klein und $0 \leq evalue < \text{infty}$
- bitscore ist besser wenn gross und $-\text{infty} < \text{bitscore} < \text{infty}$
- similarity ist besser wenn gross und $0 \leq \text{similarity} \leq 1$
- denkt darueber nach warum die bitscore in a->b und b->a unterschiedlich sein koennen