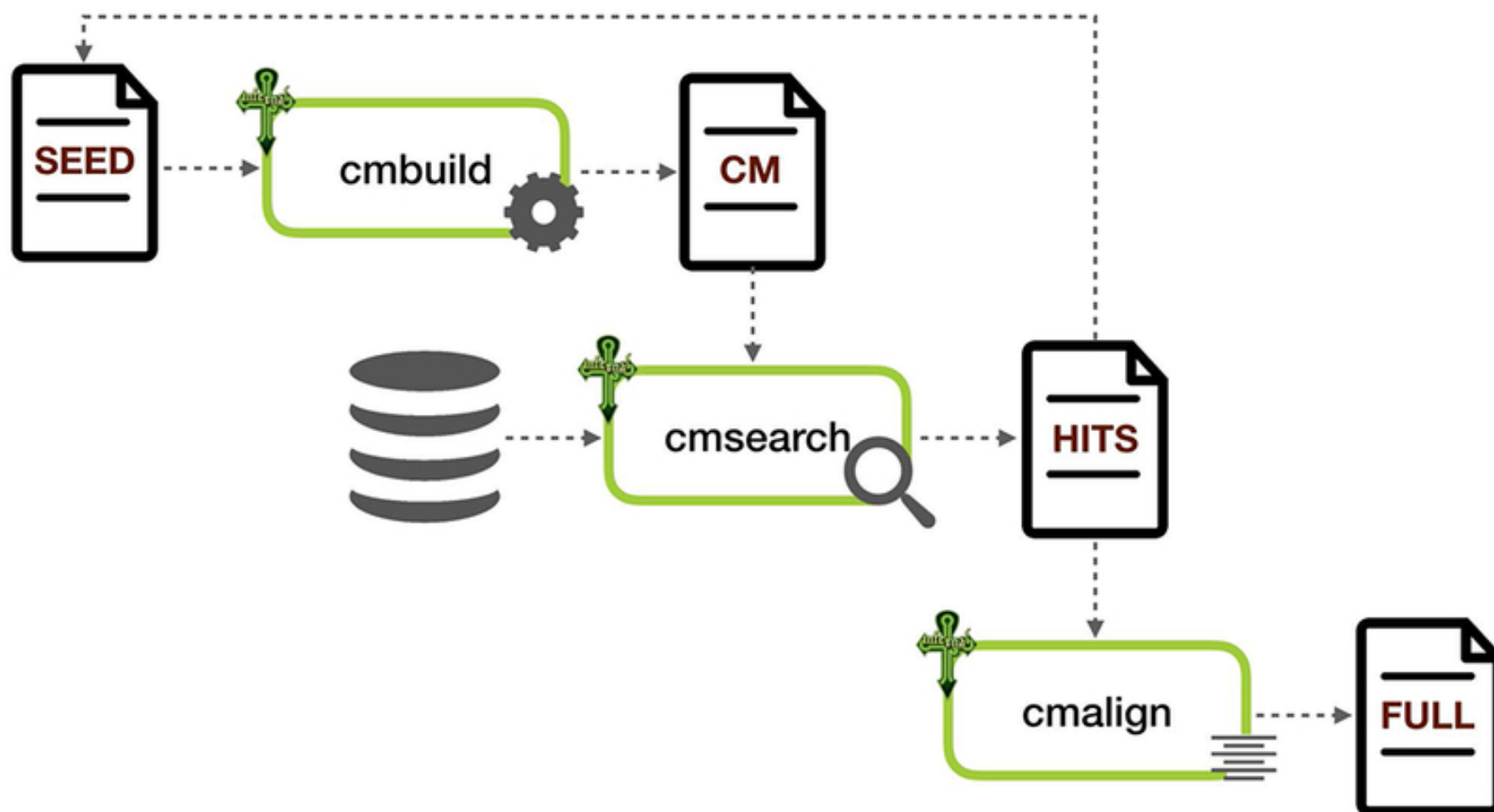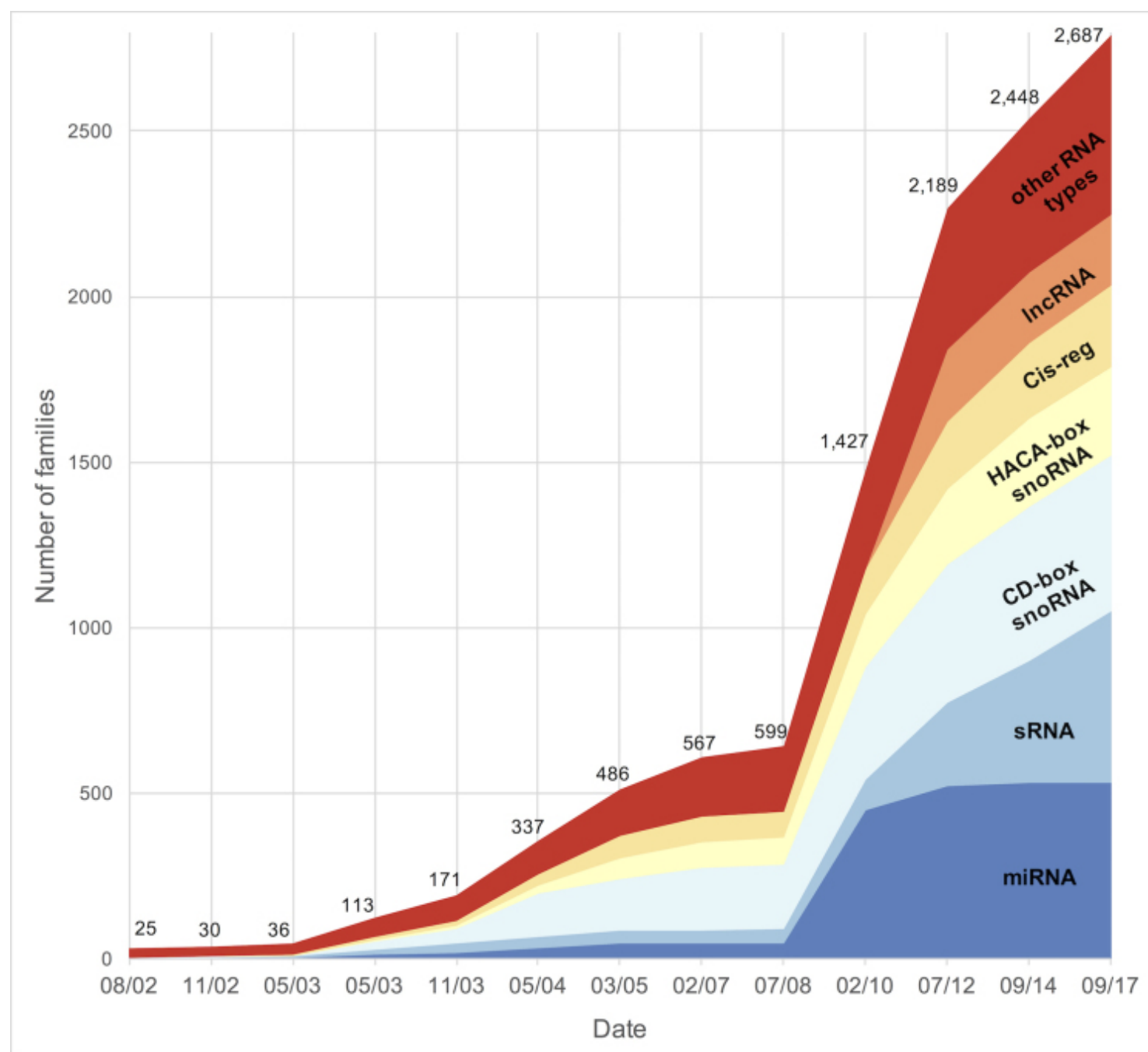# Rfam: RNA families database (3016 families)

- Each family is represented by:
  - representative SEED alignment annotated with secondary-structure
  - covariance model (CM) built from the SEED
  - hits in Rfamseq database above GA threshold (FULL)

Kalvari et al., 2018

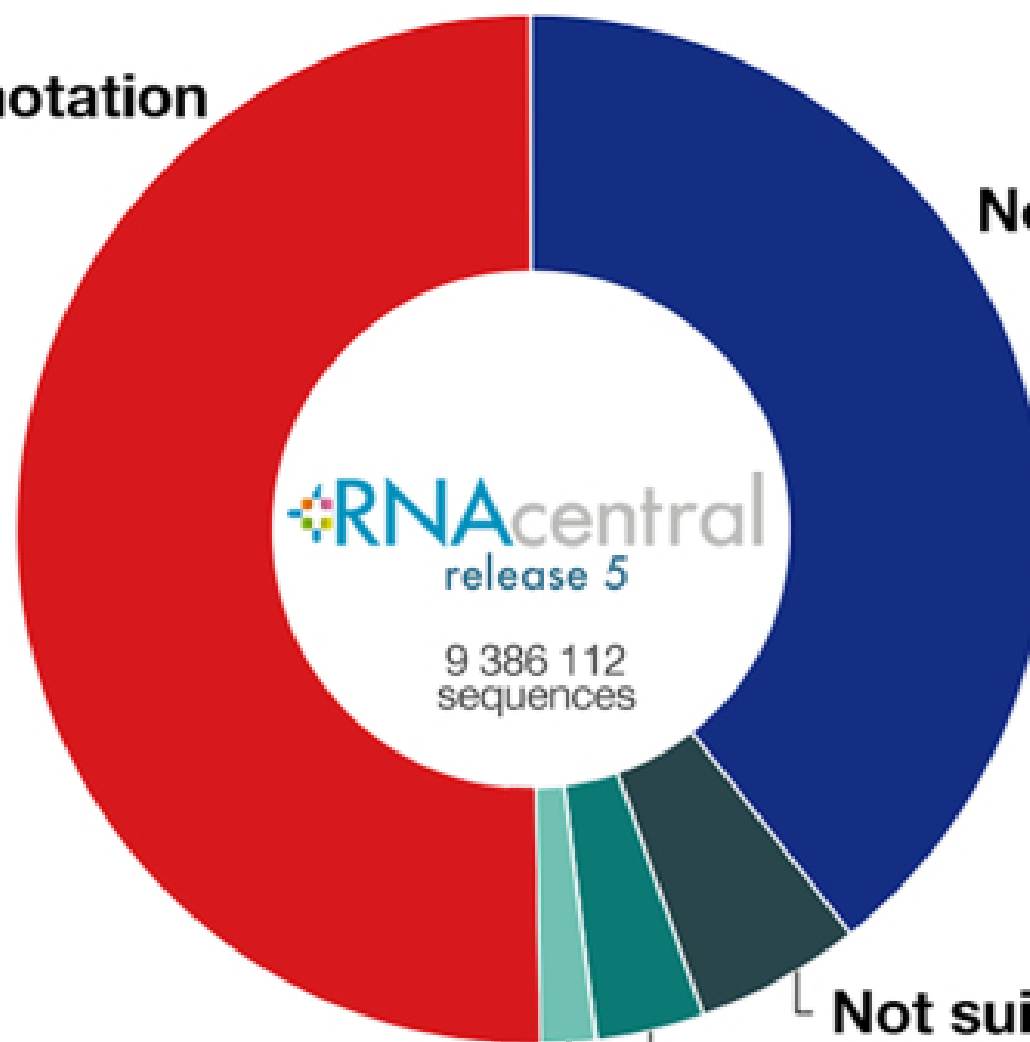With Rfam annotation
50.2%

New Rfam matches
39.4%

RNAcentral
release 5

9 386 112
sequences

Not suitable for Rfam
5.2%

Potential new Rfam families
1.8%

Undetected tRNA, rRNA
and tmRNA
3.4%

Kalvari et al., 2018

Rfamseq: switch from subset of ENA to genome-centric database

- about 8000 reference genomes

- reduces redundancy

- more scalable

Kalvari et al., 2018

Rfamseq: switch from subset of ENA to genome-centric database

- about 8000 reference genomes

- reduces redundancy

- more scalable

Rfamseq: genome-centric database means less flexible SEEDs

- previous requirement: SEED sequences must be in Rfamseq

- new approach allows any GenBank or RNAcentral sequence

- verification of sequences utilizes GenBank and RNAcentral API

Future directions for Rfam

- improve Rfam families based on crystal structures

- synchronize with mirBase

- use model reference coordinates to annotate important features

- continue to add families