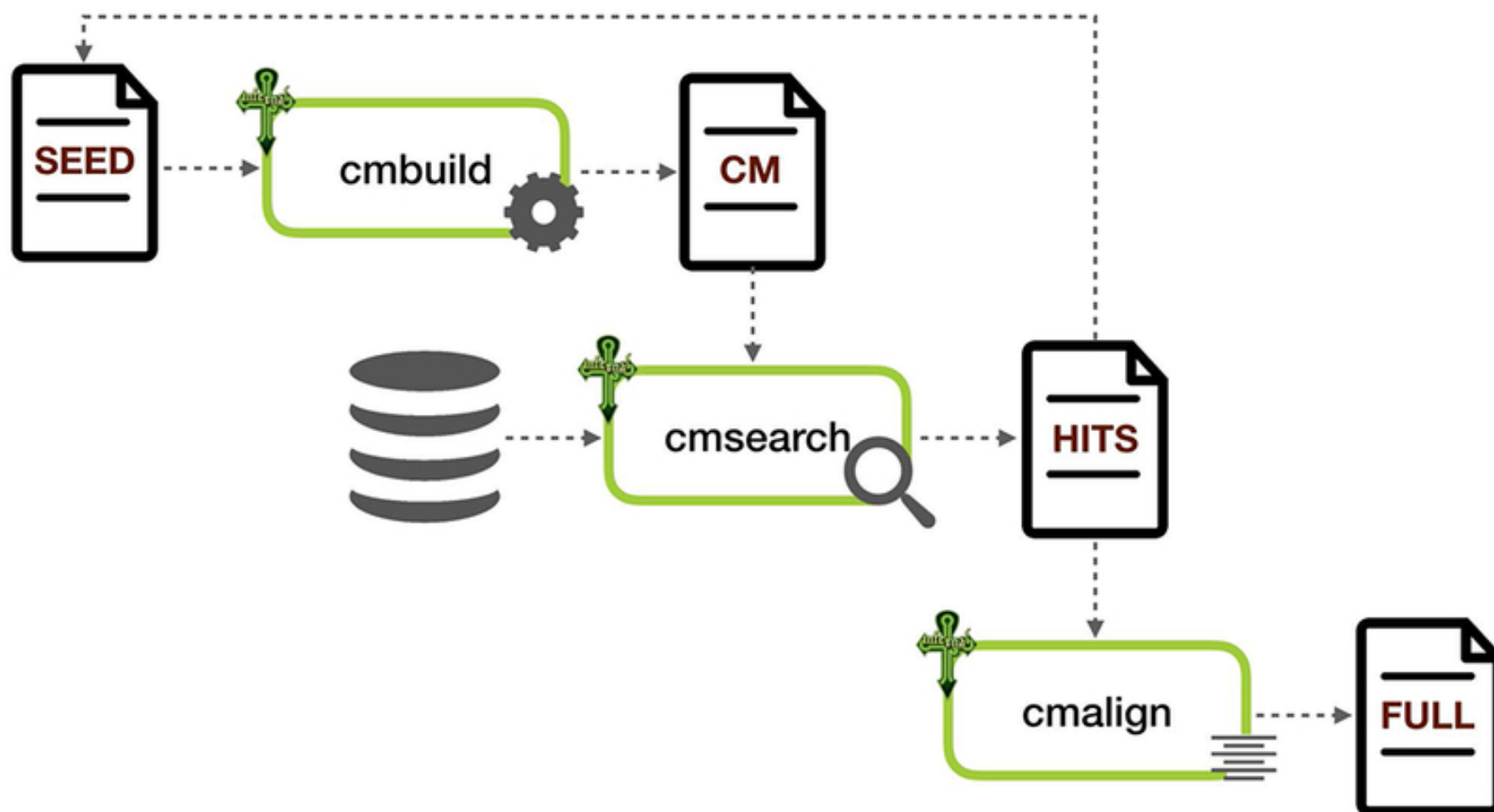
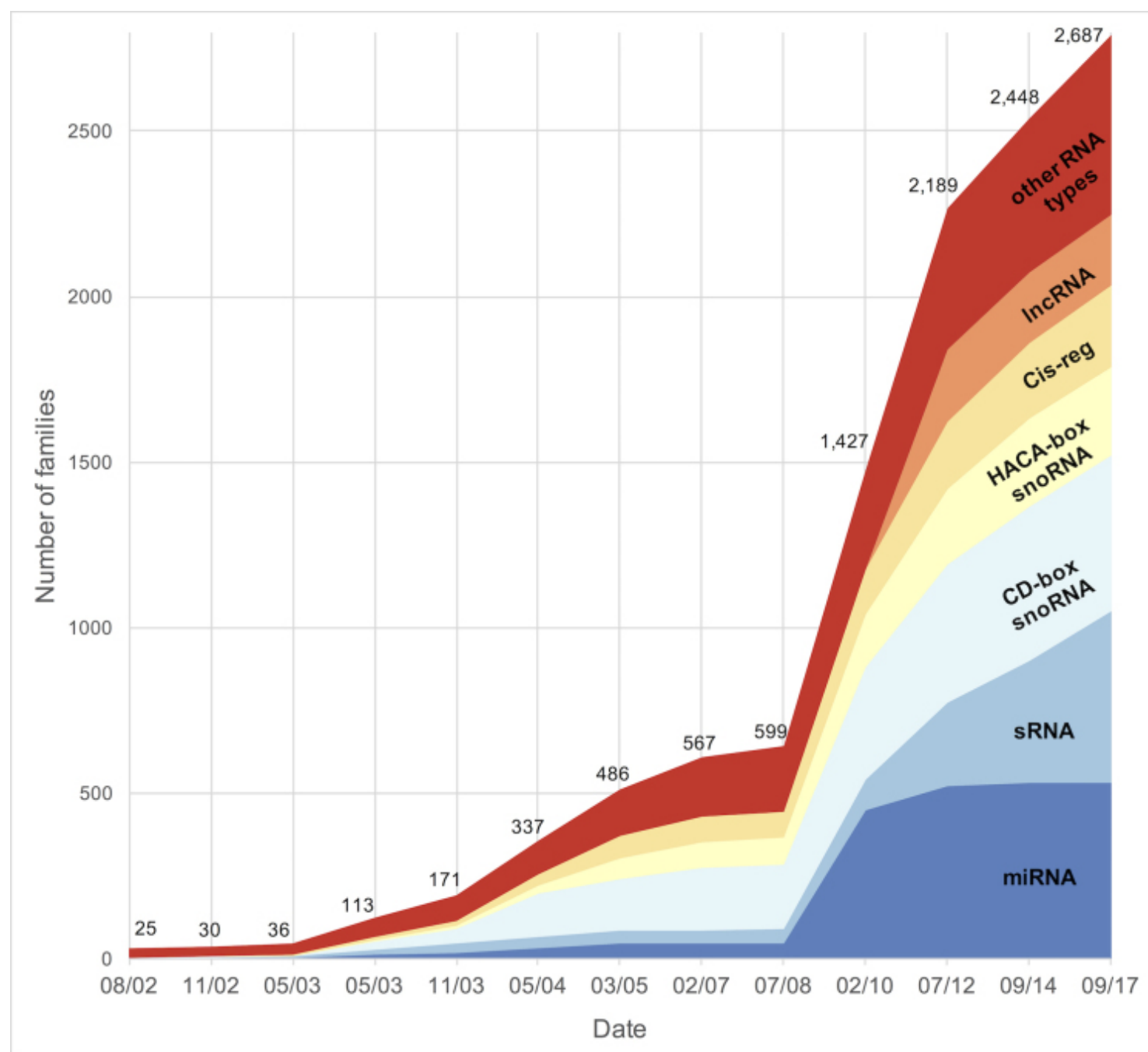


## Rfam: RNA families database (3016 families)

- 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)





**With Rfam annotation**  
50.2%

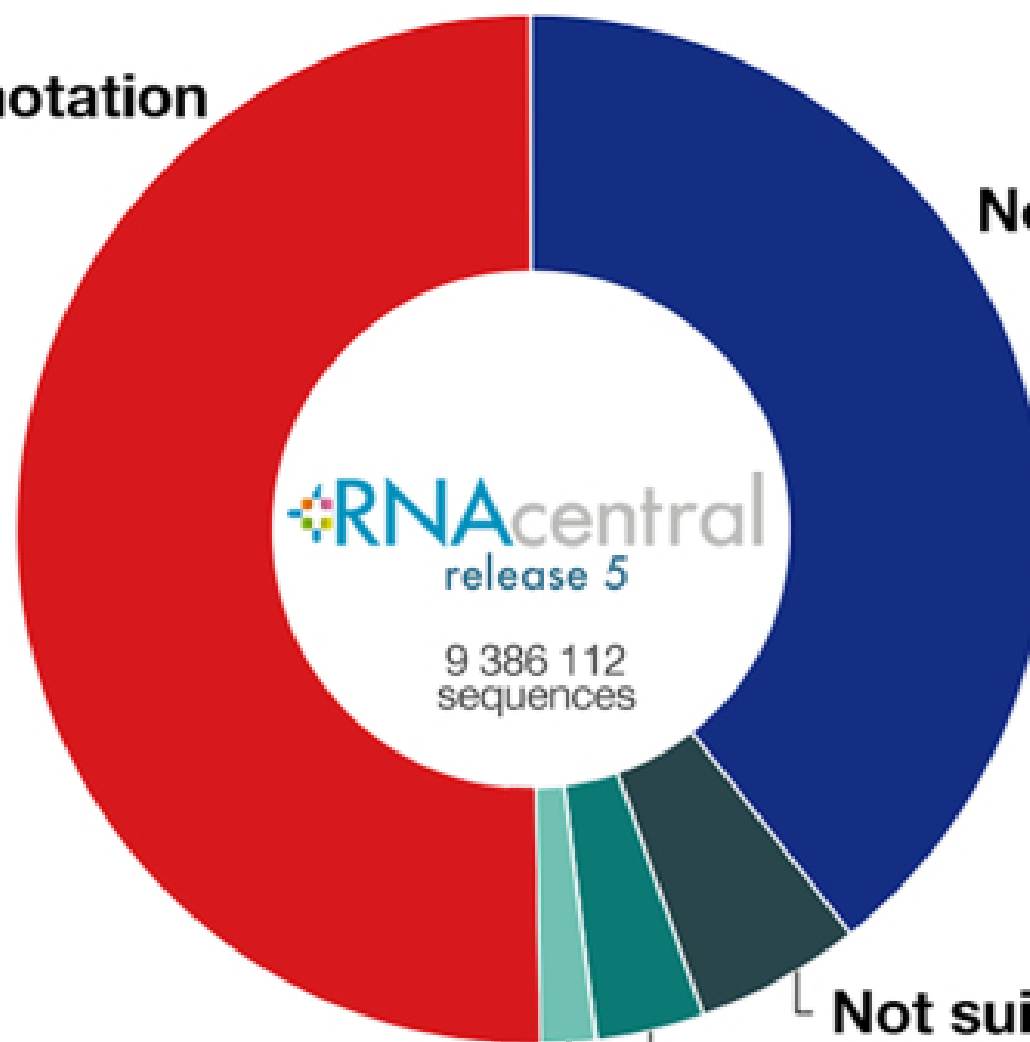
**New Rfam matches**  
39.4%

**RNAcentral**  
release 5  
9 386 112  
sequences

**Not suitable for Rfam**  
5.2%

**Undetected tRNA, rRNA  
and tmRNA**  
3.4%

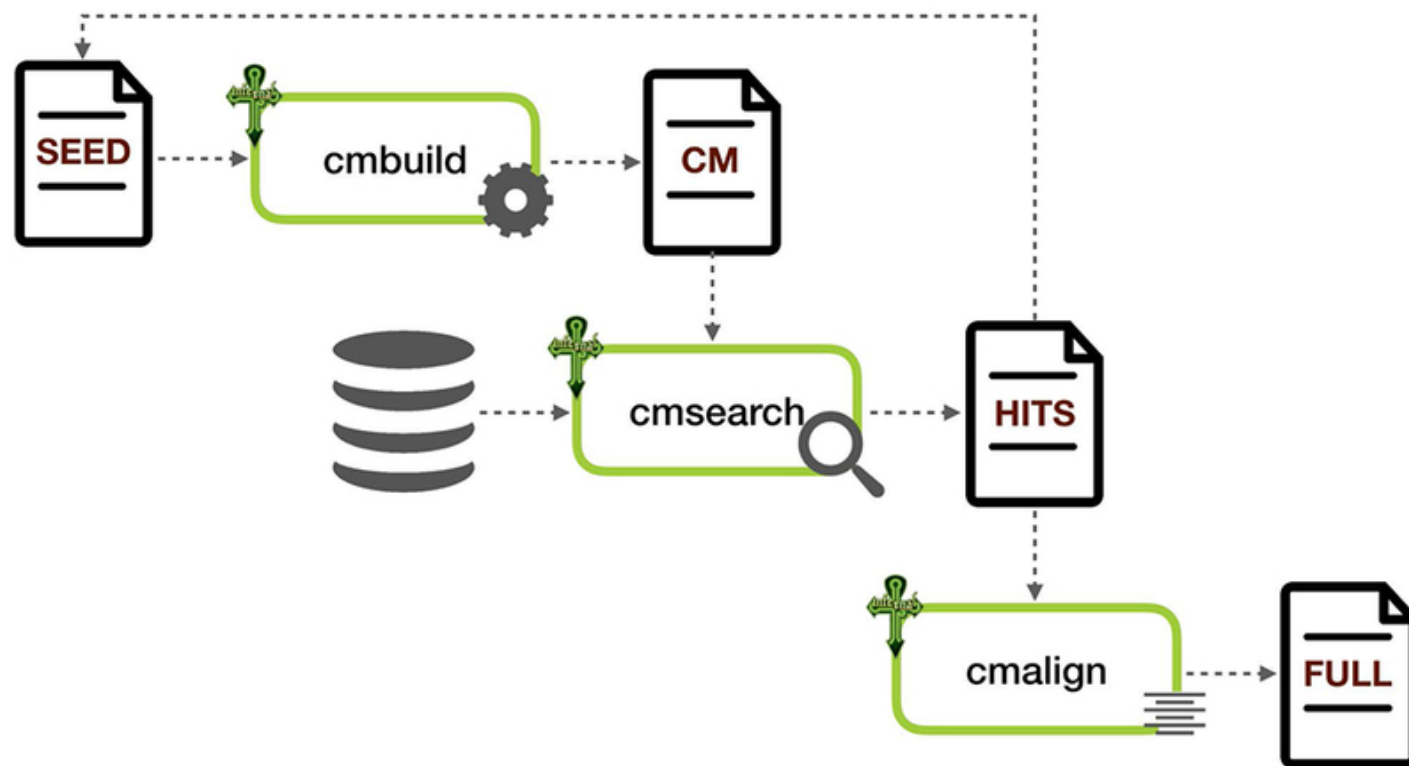
**Potential new Rfam families**  
1.8%





## Rfamseq: genome-centric database means less flexible SEEDs

- previous requirement: SEED sequences must be in Rfamseq
- new: any GenBank or RNACentral sequence can be in the SEED
- verification of sequences utilizes GenBank and RNACentral API



## Rfam in the cloud

- Ioanna has made the Rfam family building pipeline available to anyone
- Students in Daniel Gautheret's class are building Staphylococcus RNA families
- We are seeking RNA experts to improve or add to Rfam
  - new, relaxed SEED requirements make it easier to use existing curated alignments

## **Rfam in the cloud**

- Ioanna has made the Rfam family building pipeline available to anyone
- Students in Daniel Gautheret's class are building Staphylococcus RNA families
- We are seeking RNA experts to improve or add to Rfam
  - new, relaxed SEED requirements make it easier to use existing curated alignments

## **Future directions for Rfam**

- improve Rfam families based on crystal structures
- synchronize with mirBase
- viral families from Manja Marz
- use model reference coordinates to annotate important features