1. Gene family selection
   - Homolens rev.2
   - Famfetch
   - 89 families
2. Protein sequence retrieval
   - Python script
3. Enrichment of protein sequence file
   - Query-Win/Blast
   - (+30 sequences)
4. Multiple protein alignment
   - ClustalW
   - Muscle
5. Intron position identification
   - Annotate_int_mase R script using Seqinr
6. Retrieval of nucleotide sequences
   - Manual blastx for Strongylocentrotus
   - For promising gene
   - PAL2NAL
7. Choice of introns
   - Visual examination of protein alignments
     - (25 families)
   - (58 families)
   - (101 families)

EPICs defined
- 15: i1 - i17
  - (12 families)
- 11: i19 - i35
  - (3 families)
- 22: i36 - i58
  - (12 families)

*) steps 6 and 7 in reverse order in stage I