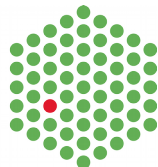# GIFTs Update

Ensembl Application Team Away Day Apr 2019

Alessandro Vullo
e! Application Team

Click to add Text

EMBL-EBI

e!Ensembl

GIFTs

Genome Integration with FuncTion and Sequence

- Connect genome/proteome worlds for max impact

- Map from genome coords to protein residues

- Main obstacle: data synchronisation

- Solution: develop common infrastructure

  – track entities btw resources

  – synchronise content

  – share curational tasks

  – support mapping btw entities at nucleotide/AA precision

Goals

- Support:
  - DNA features ↔ proteins projections
  - common curation/analysis shared across resources
- System for agreed reference gene/transcript/protein sets from e! and UniProt
- Resource for genome integration with function and sequence

Deliverables

- GIFTs database (e!/Uniprot)
  - known genes, transcripts, proteins and their relationships
  - all previous transcript versions stored with relationships to genes
  - drives the mapping of UniProtKB sequences to genomes
  - support xrefs management and share curation interface

- Curation interface
  - view mappings
  - validate/edit automatic mappings

Deliverables (2)

- Protein-Genome Mapping Pipeline
  - DB remains in synch with e! and UniProt releases
  - automatically align proteins to reference genomes
    - store matches
    - <100% matches trigger further analysis and/or resolution through curation interface

- GIFTs Resource
  - present DB mappings as simple EMBL-EBI webpage
  - download of mapping files
  - project docs and use cases

EMBL-EBI

Milestones

- Year 1 (2017)
  - DB schema designed, implementated, populated
  - API prototype to support protein-genome mapping pipeline
  - pipeline developed (use API and GIFTs DB)
  - commence work on curation interface

- Year 2 (2018)
  - develop curation interface (use API and GIFTs DB)
  - develop GIFTs website
  - develop automated method to query/combine e!/UniProt annotation → agreed reference genome/protein sets

Timeline

- May '18 (6 months to go!)
  - prototype schema designed
  - DB test server instance
  - *API backend skeleton (Django), no endpoints*
  - genebuild pipeline initial implementation (Perl)
  - front end work starts, mockup using precompiled JSON
- June '18
  - Web VM/virtual users request
  - several schema redesigns
  - back end API
    - prototype endpoint for bulk load gene/transcripts
    - endpoints to support front end build

Timeline

- July '18
  - new schema, more API endpoints (pipeline/front end)
  - loaded human/mouse
  - virtual users, no VMs yet!
    - plan B: OpenStack (can't reach DB, mirror)
  - 1st prototype interface
- August '18
  - several schema/front end API updates
  - 1st demo curation interface
  - finally, the VMs!
- September '18
  - several schema/front end API updates
  - first instances running dev/prod (HX/HH)
    - semi-automated deployment process
  - production DB VM request (HX/HH)
  - Matt leaves!

EMBL-EBI

Timeline

- October '18
  - production DB VMs ready (HX)
  - pipeline tests alignment API
  - feedback on curation tool

- November '18
  - new requirement: unmapped entries
  - schema/API updates
  - authentication set up/tests
  - dev nightly builds
  - DB server monitoring/tuning

Timeline

- December '18
  - debugging web-prod URL mapping, Apache rewrite rules
  - tuning Apache (multi-process)
  - configurable deployment
  - pre-release out: http://www-test.ebi.ac.uk/gifts
    - bug-fixes
    - collaborative workflow simulation starts
- January '19
  - curators feedback
  - search issues (data/speed), bug-fixes
  - schema/API improvements
  - new requirement: mapped with unmapped

Timeline

- February '19
  - GIFTs demo at EBI service retreat
  - release/workflow process (Beth)
  - schema/API/front end updates
    - mapped with unmapped
    - search speed improvements
  - Load Balancer problems
  - VM kernel upgrade: build process disruption
- March '19
  - GIFTs presented to UniProt collaborators
  - feature requests:
    - review unmapped
    - changes to existing endpoints
  - REST based e! pipeline time out:
    - REST polling (task queue manager)
  - Mark steps in

EMBL-EBI

e!

To Be Completed

- API docs

- REST polling
  - Celery/Redis deployed

- review unmapped (front end)

EMBL-EBI