KBase
PREDICTIVE BIOLOGY

DOE Systems Biology Knowledgebase

INTEGRATION
AND MODELING
for PREDICTIVE
BIOLOGY

Software Carpentry Workshop

July 17-18
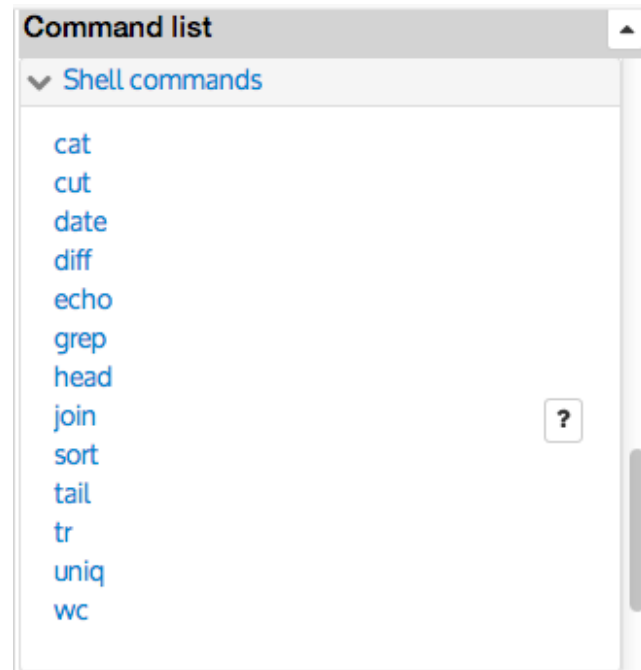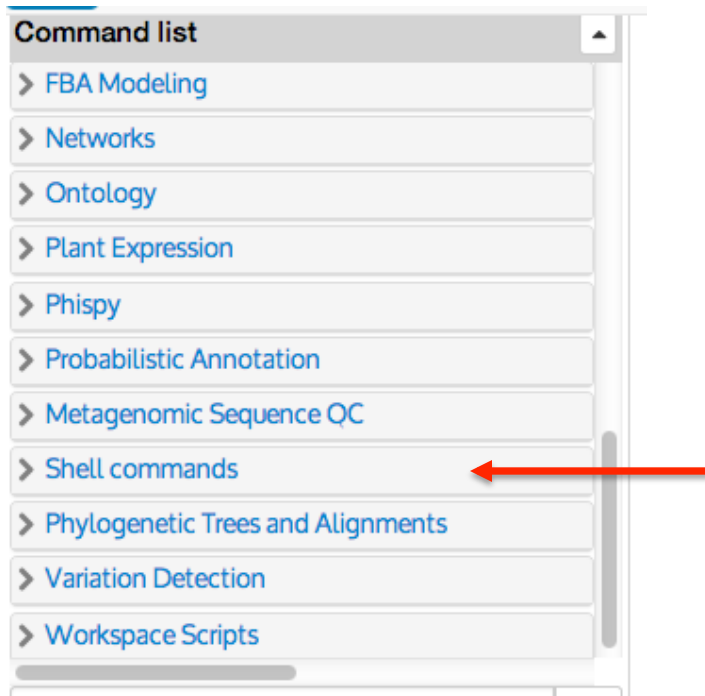
Portland, OR

KBase-Plants Demo

- **Introduction to IRIS**
  - File upload/download
  - Central Store
  - Ontology
  - Expression
- **Introduction to Narrative**
  - Central Store

IRIS

http://iris.kbase.us

- **Shell commands**

  - Hint: You can enlarge font size in the browser to see text better

  - Click on 'Shell commands' in side-bar

Office of Biological and
Environmental Research

U.S. DEPARTMENT OF
ENERGY
Office of Science

KBase
PREDICTIVE BIOLOGY
DOE Systems Biology Knowledgebase

IRIS

http://iris.kbase.us

- **Redirection and File upload/download**
  - > echo 'Hello World!' > Hello_World.txt
  - Controls allow files to be uploaded/downloaded/edited

No processes running

File Browser

🗋 Hello_World.txt

```
>/ echo "Hello World!" > Hello_World.txt
Command Completed
```

IRIS

http://iris.kbase.us

- **Central Store**

  > all_entities_Genome –help

  > all_entities_Genome | wc

  > all_entities_Genome --fields source_id,scientific_name,domain,pegs| grep Arabidopsis

```
>/ all_entities_Genome | wc
  13111    13111   139482
Command Completed
```

```
>/ all_entities_Genome --fields source_id,scientific_name,domain,pegs | grep Arabidopsis
kb|g.140055    Alyrata.v.1.0      Arabidopsis lyrata subsp. lyrata    Eukaryota    32667
kb|g.3899      Athaliana.TAIR10   Arabidopsis thaliana                Eukaryota    35386
Command Completed
```

IRIS

http://iris.kbase.us

- **Relationships (Source)**
  - Two sources of Plant Genomes: Phytozome and EnsemblPlant

```
>/ echo "kb|g.3899" | get_relationship_WasSubmittedBy --to id
kb|g.3899     EnsemblPlant
Command Completed
```

```
>/ echo "EnsemblPlant" | get_relationship_Submitted --to scientific_name,source_id
EnsemblPlant     Arabidopsis lyrata subsp. lyrata     Alyrata.v.1.0
EnsemblPlant     Aegilops tauschii                    Atauschii.ASM34733v1
EnsemblPlant     Cyanidioschyzon merolae strain 10D   Cmerolae.ASM9120v1
EnsemblPlant     Chlamydomonas reinhardtii            Creinhardtii.v3.0
EnsemblPlant     Vitis vinifera                       Vvinifera.IGGP_12x
EnsemblPlant     Selaginella moellendorffii           Smoellendorffii.v1.0
```

```
>/ echo "kb|g.3899" | get_relationship_WasSubmittedBy --to id | cut -f2 | get_relationship_Submitted --to scientific_name,source_id
```

IRIS

http://iris.kbase.us

- **Relationships (Feature)**

  - Three feature types: locus, mRNA, and CDS

```
>/ echo "kb|g.3899" | get_relationship_IsOwnerOf | wc
 110659   110659 1327908
Command Completed


>/ echo "kb|g.3899" | get_relationship_IsOwnerOf --to id,source_id,feature_type | head
kb|g.3899      kb|g.3899.CDS.35386    ATMG00590.1.CDS    CDS
kb|g.3899      kb|g.3899.CDS.35387    ATMG00480.1.CDS    CDS
kb|g.3899      kb|g.3899.CDS.35388    ATCG00830.1.CDS    CDS
kb|g.3899      kb|g.3899.CDS.35389    ATMG00090.1.CDS    CDS
kb|g.3899      kb|g.3899.CDS.35390    ATMG01100.1.CDS    CDS
kb|g.3899      kb|g.3899.CDS.35391    AT5G01030.2.CDS    CDS
kb|g.3899      kb|g.3899.CDS.35392    ATMG00110.1.CDS    CDS
kb|g.3899      kb|g.3899.CDS.35393    ATCG00870.1.CDS    CDS
kb|g.3899      kb|g.3899.CDS.35394    AT5G01390.2.CDS    CDS
kb|g.3899      kb|g.3899.CDS.35395    ATMG01140.1.CDS    CDS
Command Completed

>/ echo "kb|g.3899" | get_relationship_IsOwnerOf --to id,source_id,feature_type | cut -f4 | sort | uniq -c
   35386 CDS
   33602 locus
   41671 mRNA
Command Completed
```

IRIS

http://iris.kbase.us

- **Sequences**
  - fids_to_protein_sequences

```
>/ echo "kb|g.3899" | get_relationship_IsOwnerOf --to id,source_id,feature_type | grep CDS 2> /dev/null | cut -f2 | head | fids_to_protein_sequences
>kb|g.3899.CDS.35386
MTIRNQRFSLLKQPISSTLNQHLVDYPTPSNLSYWWGFGPLAGTMILSVLSSPALVSGLMVARAKNLVHSVLFPIPIFFSINQLFHYFCRLPIIKHLATKCQLLLFLISHFLLLLVLTKLVLDLGGYLFMDDLSRALSQFVPGFSGGLGG
>kb|g.3899.CDS.35387
MPQLDKFTYFSQFFWLCLFFFTFYIFICNDGDGVLGISRILKLRNQLLSHRGKTIRSKDPNSLEDLLRKGFSTGVSYMYASLFEVSQWCKAVDLLGKRRKITLISCFGEISGSRGMERNILYNISKSSPSNTGRWITCRNCRNDIMLIHV
>kb|g.3899.CDS.35388
MAIHLYKTSTPSTRNGAVDSQVKSNPRNNLICGQHHCGKGRNARGIITARHRGGGHKRLYRKIDFRRNAKDIYGRIVTIEYDPNRNAYICLIHYGDGEKRYILHPRGAIIGDTIVSGTEVPIKMGNALPLTDMPLGTAIHNIEITLGRGG
>kb|g.3899.CDS.35389
MARKGNPISVRLGKNRSSDSSRFSEYYYGKFVYQDVNLRSYFGSIRPPTRLTFGFRLGRCIILHFPKRTFIHFFLPRRPRRLKRREKTRPGKEKGRWWTTFGKAGPIECLHSSDDTEEERNEVRGRGARKRVESIRLDDRKKQNEIRGWP
>kb|g.3899.CDS.35390
MQSPAMKRIKSSSHSRWDGSGSVNEMPFPSTIRLQGSFWECSTRRHMCYILRYLFRANGHRHFSYERLDCRNQTLRLPDHLYQPSRPHLLPHLSQLLLVRDSGYL
>kb|g.3899.CDS.35391
MAEKEGKASGTDNDNRVKTKRSSHSRRKECVNKSLEHNDELVKYMSKLPGYLQRIERGEESVHQSNVLNVGVLDWESLQRWKHGRAKGGEISGRSERKVSTIATTSTSGVVVPNDSANRCKIDDQVHTCSNLGKVKASRDLQYSLEPQLA
>kb|g.3899.CDS.35392
MRRLFLELYHKLIFSSTPITSFSSFLSYIVVTPLMLGFEKDFSCHSHLGPIRIPPLFPFPPAPFPRNEKEDGTLELYYLSTYCLPKILLLQLVGHRVIQISRVFCGFPMLQLSYQFGRSGMDRLNIPLGSLVLTLLCGIHSRSALGITSS
>kb|g.3899.CDS.35393
MLLLKHGRIEILDQNTMYGWYELPKQEFLNSEQPELLLTTSKKFPLIKDGNPLENQKYACRIKLLLLSVPITNQLNN
```

IRIS

http://iris.kbase.us

- **Gene Ontology Service**

  - get_goidlist

```
>/ echo "kb|g.3899" | get_relationship_IsOwnerOf --to id,source_id,feature_type | grep locus 2> /dev/null | head | cut -f2 | get_goidlist
kb|g.3899.locus.100      GO:0005634    cellular_component    IEA    nucleus
kb|g.3899.locus.100      GO:0003677    molecular_function    ISS    DNA binding
kb|g.3899.locus.100      GO:0006355    biological_process    IEA    regulation of transcription, DNA-dependent
kb|g.3899.locus.100      GO:0003700    molecular_function    ISS    sequence-specific DNA binding transcription factor activity
kb|g.3899.locus.10       GO:0006333    biological_process    IEA    chromatin assembly or disassembly
kb|g.3899.locus.10       GO:0005634    cellular_component    IEA    nucleus
kb|g.3899.locus.10       GO:0000785    cellular_component    IEA    chromatin
kb|g.3899.locus.10       GO:0003682    molecular_function    IEA    chromatin binding
kb|g.3899.locus.10       GO:0003676    molecular_function    IEA    nucleic acid binding
kb|g.3899.locus.10000    GO:0005777    cellular_component    IDA    peroxisome
kb|g.3899.locus.10000    GO:0003924    molecular_function    IEA    GTPase activity
kb|g.3899.locus.10000    GO:0009791    biological_process    IGI    post-embryonic development
kb|g.3899.locus.10000    GO:0005739    cellular_component    IDA    mitochondrion
kb|g.3899.locus.10000    GO:0005525    molecular_function    IEA    GTP binding
kb|g.3899.locus.10000    GO:0005886    cellular_component    IDA    plasma membrane
```

IRIS

http://iris.kbase.us

- **Plant Expression Service (Plant Ontology and Environment Ontology)**
  - get_all_po
  - get_po_sampleidlist

```
>/ get_all_po | wc
    59      183     1500
Command Completed
```

```
>/ get_all_po | grep leaf
PO:0009025    leaf
PO:0000014    rosette leaf
PO:0001054    4 leaf senescence stage
PO:0006339    juvenile leaf
PO:0020137    leaf apex
PO:0020040    leaf base
PO:0000013    cauline leaf
Command Completed
```

```
>/ get_all_po | grep leaf | cut -f1 | get_po_sampleidlist | head
PO:0000014    Arabidopsis thaliana    GSM128778
PO:0000014    Arabidopsis thaliana    GSM128779
PO:0000014    Arabidopsis thaliana    GSM128780
PO:0000014    Arabidopsis thaliana    GSM131060
PO:0000014    Arabidopsis thaliana    GSM131061
PO:0000014    Arabidopsis thaliana    GSM131062
PO:0000014    Arabidopsis thaliana    GSM131063
PO:0000014    Arabidopsis thaliana    GSM131064
PO:0000014    Arabidopsis thaliana    GSM131065
PO:0000014    Arabidopsis thaliana    GSM131501
Command Completed
```

Office of Biological and
Environmental Research

U.S. DEPARTMENT OF
ENERGY
Office of Science

KBase
PREDICTIVE BIOLOGY
DOE Systems Biology Knowledgebase

IRIS

http://iris.kbase.us

- **Plant Expression Service (Plant Ontology and Environment Ontology)**
  - get_experiments_by_sampleid

```
>/ echo "GSM128778 " | get_experiments_by_sampleid | head

expression table for GSM128778:
"","GSM128778"
"kb|g.3899.locus.11535",6.83705385166372
"kb|g.3899.locus.11338",7.48470915886701
"kb|g.3899.locus.11424",7.39551586205802
"kb|g.3899.locus.11419",3.46092768149346
"kb|g.3899.locus.11367",6.02223356219287
"kb|g.3899.locus.11298",4.53824306449828
"kb|g.3899.locus.11377",5.45711806164956
Command Completed
```

```
>/ echo "PO:0000014" | get_po_sampleidlist --species=Ath | cut -f3 | head | get_experiments_by_sampleid > PO00000014_GEO_Ath.txt
Command Completed
```

Office of Biological and
Environmental Research

U.S. DEPARTMENT OF
ENERGY
Office of Science

KBase
PREDICTIVE BIOLOGY
DOE Systems Biology Knowledgebase

Narrative

http://demo.kbase.us

My Projects & Narratives ⓘ

+New Project	☑ Owned by me ☐ Shared with me

Search: [_____]

Name ⇅	Last Modified ⇅		
▾ Project **home** - Me	Shared with: *	**+ Narrative**	Data | Manage
auto5	May 22, 2014		
▾ Project **Plant_Genomes** - Me	Shared with: *	**+ Narrative**	Data | Manage
List_Plant_Genomes	June 12, 2014		
▾ Project **Software_Carpentry** - Me		**+ Narrative**	Data | Manage
Test	6 hours ago		

← Previous   **1**   Next →

Narrative

http://demo.kbase.us

```
In [2]: pht = %inv_run echo 'Phytozome' | get_relationship_Submitted --to id,scientific_name,source_id,pegs
```

```
In [3]: for x in pht.split('\n'):
            x=x.split('\t')
            if(len(x)!=5):
                continue
            print '{0:20}{1:20}{2:40}{3:40}{4:6}'.format(*x).strip()
```

```
Phytozome           kb|g.166803         Micromonas pusilla RCC299             Mpusilla_RCC299.JGI1.0              10103
Phytozome           kb|g.166804         Aquilegia coerulea Goldsmith          Acoerulea.JGI1.1                   41063
Phytozome           kb|g.166805         Citrus clementina                     Cclementina.JGI1.0                 33929
Phytozome           kb|g.166806         Carica papaya                         Cpapaya.ASGPB0.4-Dec-2             27793
Phytozome           kb|g.166807         Chlamydomonas reinhardtii             Creinhardtii.v5.3.1                19526
Phytozome           kb|g.166808         Capsella rubella -JGI v1.0            Crubella.Capsella-rubella--JGI1.0  28447
Phytozome           kb|g.166809         Cucumis sativus                       Csativus.ROche-JGI1                30364
Phytozome           kb|g.166810         Citrus sinensis                       Csinensis.JGI1.1                   46147
Phytozome           kb|g.166811         Coccomyxa subellipsoidea C-169        Csubellipsoidea_C-169.JGI2.0       9629
Phytozome           kb|g.166812         Eucalyptus grandis                    Egrandis.JGI1.1                    46315
Phytozome           kb|g.166813         Fragaria vesca                        Fvesca.v1.1                        32831
Phytozome           kb|g.166814         Gossypium raimondii                   Graimondii.JGI2.1                  77267
Phytozome           kb|g.166815         Linum usitatissimum                   Lusitatissimum.BGIv1.0             43484
Phytozome           kb|g.166816         Malus domestica                       Mdomestica.GDRv1.0                 63517
Phytozome           kb|g.166817         Manihot esculenta                     Mesculenta.JGI4.1                  34151
Phytozome           kb|g.166818         Mimulus guttatus                      Mguttatus.JGI-IM621.1              28282
Phytozome           kb|g.166819         Ostreococcus lucimarinus              Olucimarinus.JGI1.0                7796
Phytozome           kb|g.166820         Prunus persica                        Ppersica.JGI1.0                    28701
Phytozome           kb|g.166821         Panicum virgatum                      Pvirgatum.JGI0.0                   70071
Phytozome           kb|g.166822         Phaseolus vulgaris                    Pvulgaris.JGI1.0                   31638
Phytozome           kb|g.166823         Ricinus communis                      Rcommunis.TIGR-Release0.1          31221
Phytozome           kb|g.166828         Populus trichocarpa                   Ptrichocarpa.JGI3.0                73013
```