

Expression Divergence of Chemosensory Genes between *Drosophila sechellia* and Its Sibling Species and Its Implications for Host Shift

報告人：

108753102 資科碩一 吳映函

108753109 資科碩一 陳庭軒

Meng-Shin Shiao, Jia-Ming Chang,
Wen-Lang Fan, Mei-Yeh Jade Lu,
Cedric Notredame, Shu Fang,
Rumi Kondo, and Wen-Hsiung Li

Abstract

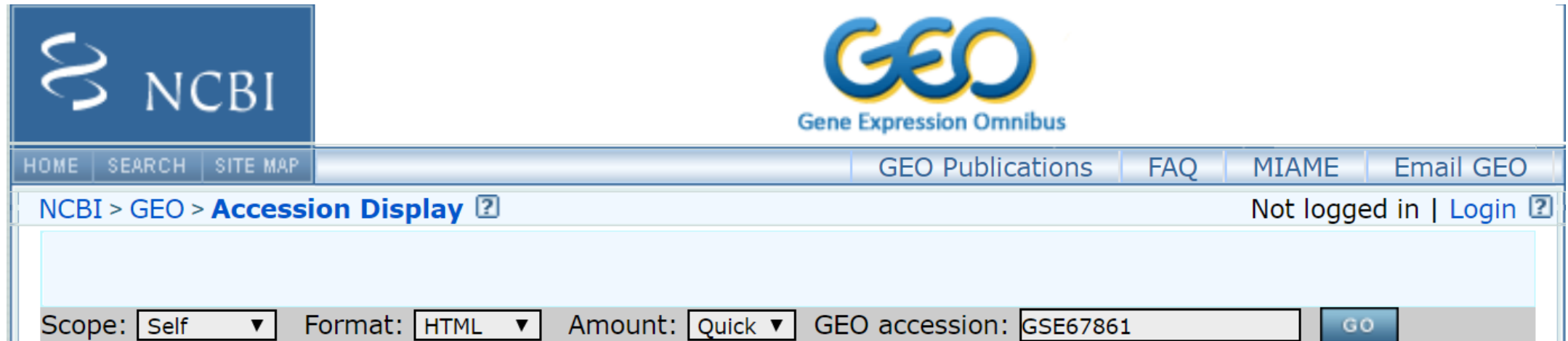
- We wanted to know whether gene expression, particularly chemosensory genes, has diverged between *D. sechellia* and its two sibling species.
- we found a higher percentage of chemosensory genes differentially expressed in the *D. sechellia* lineage(7.8%) than in the *D. simulans* lineage (5.4%)
- Our observations suggest that the host shift of *D. sechellia* was associated with the enrichment of differentially expressed, particularly upregulated, chemosensory genes.

Materials and Methods

- We chose the most closely related generalist species, **D. simulans**, as a comparison and *D. melanogaster* as the outgroup generalist species.
- From *D. melanogaster* 3 to **D. sechellia** and *D. mauritiana*.
- The orthologous gene set (OGS) was manually curated by excluding genes duplicated in any of the species and only single copy genes in all three species were selected.
- Or22a has two distinct gene IDs in *D. simulans* (FBgn0068650 and FBgn0194492) and in *D. sechellia* (FBgn0171736 and FBgn0259897), whereas Or22b has two FlyBase IDs in *D. simulans* (FBgn0068649 and FBgn0194493).

Dataset

- NCBI GEO
 - GSE67587
 - GSE67861
 - GSE67862



The screenshot shows the NCBI GEO website interface. At the top left is the NCBI logo. To its right is the GEO logo with the text "Gene Expression Omnibus". Below these are navigation links: HOME, SEARCH, SITE MAP, GEO Publications, FAQ, MIAME, and Email GEO. A breadcrumb trail reads "NCBI > GEO > Accession Display". On the right, it says "Not logged in | Login". The main content area is a light blue box. At the bottom, there is a search bar with the following fields: "Scope:" with a dropdown menu set to "Self", "Format:" with a dropdown menu set to "HTML", "Amount:" with a dropdown menu set to "Quick", and "GEO accession:" with a text input field containing "GSE67861". A "GO" button is located to the right of the accession number field.

NCBI

GEO
Gene Expression Omnibus

HOME SEARCH SITE MAP GEO Publications FAQ MIAME Email GEO

NCBI > GEO > **Accession Display** ?

Not logged in | Login ?


Scope: Self ▼ Format: HTML ▼ Amount: Quick ▼ GEO accession: GSE67861 GO

Data Accessibility

- GSE67587 for *D. sechellia* (Taiwan)
- GSE67861 for *D. sechellia* (Japan)
- GSE67862 for *D. simulans* (Japan)


Dataset

Samples (6)

 [Less...](#)


[GSM1650068](#) Dsec male TW rep1
[GSM1650069](#) Dsec male TW rep2
[GSM1650070](#) Dsec male TW rep3
[GSM1650071](#) Dsec female TW rep1
[GSM1650072](#) Dsec female TW rep2
[GSM1650073](#) Dsec female TW rep3

Samples (6)

 [Less...](#)

[GSM1657327](#) Dsec male JP rep1
[GSM1657328](#) Dsec male JP rep2
[GSM1657329](#) Dsec male JP rep3
[GSM1657330](#) Dsec female JP rep1
[GSM1657331](#) Dsec female JP rep2
[GSM1657332](#) Dsec female JP rep3

Samples (6)

 [Less...](#)

[GSM1657333](#) Dsim male JP rep1
[GSM1657334](#) Dsim male JP rep2
[GSM1657335](#) Dsim male JP rep3
[GSM1657336](#) Dsim female JP rep1
[GSM1657337](#) Dsim female JP rep2
[GSM1657338](#) Dsim female JP rep3

GSE67587 for *D. sechellia* (Taiwan)

Samples	Run	# of Bases(G)	Size(Gb)
GSM1650068 Dsec male TW rep1	SRR1952772	6.4G	4.1Gb
GSM1650069 Dsec male TW rep2	SRR1952773	5.9G	3.7Gb
GSM1650070 Dsec male TW rep3	SRR1952774	6.2G	3.9Gb
GSM1650071 Dsec female TW rep1	SRR1952775	7.3G	4.6Gb
GSM1650072 Dsec female TW rep2	SRR1952776	6.7G	4.2Gb
GSM1650073 Dsec female TW rep3	SRR1952777	7.0G	4.4Gb

GSE67861 for *D. sechellia* (Japan)

Samples	Run	# of Bases(G)	Size(Gb)
GSM1657327 Dsec male JP rep1	SRR1973486	8.1G	5.2Gb
GSM1657328 Dsec male JP rep2	SRR1973487	7.5G	4.7Gb
GSM1657329 Dsec male JP rep3	SRR1973488	7.8G	4.9Gb
GSM1657330 Dsec female JP rep1	SRR1973489	6.1G	3.9Gb
GSM1657331 Dsec female JP rep2	SRR1973490	5.8G	3.6Gb
GSM1657332 Dsec female JP rep3	SRR1973491	6.0G	3.8Gb

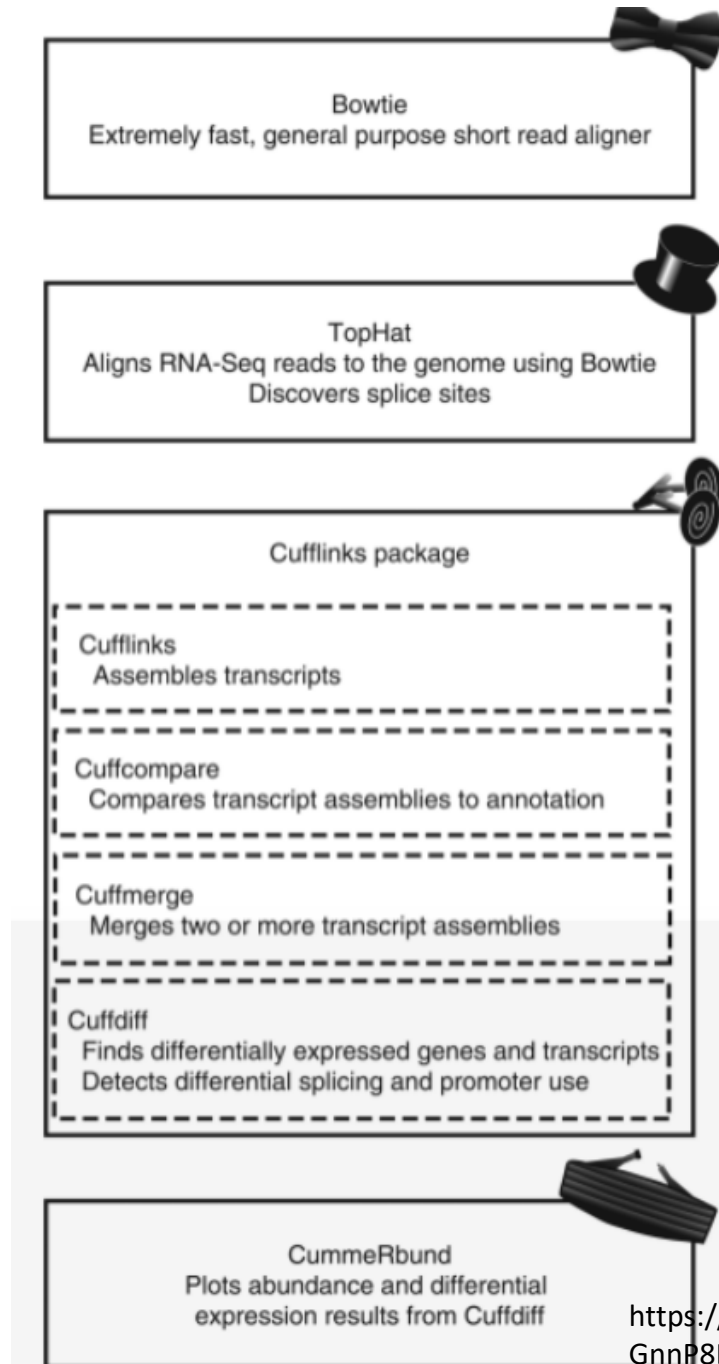
GSE67862 for D. simulans (Japan)

Samples	Run	# of Bases(G)	Size(Gb)
GSM1657333 Dsim male JP rep1	SRR1973492	5.7G	3.6Gb
GSM1657334 Dsim male JP rep2	SRR1973493	5.3G	3.3Gb
GSM1657335 Dsim male JP rep3	SRR1973494	5.5G	3.5Gb
GSM1657336 Dsim female JP rep1	SRR1973495	7.7G	5.0Gb
GSM1657337 Dsim female JP rep2	SRR1973496	7.1G	4.5Gb
GSM1657338 Dsim female JP rep3	SRR1973497	7.4G	4.7Gb

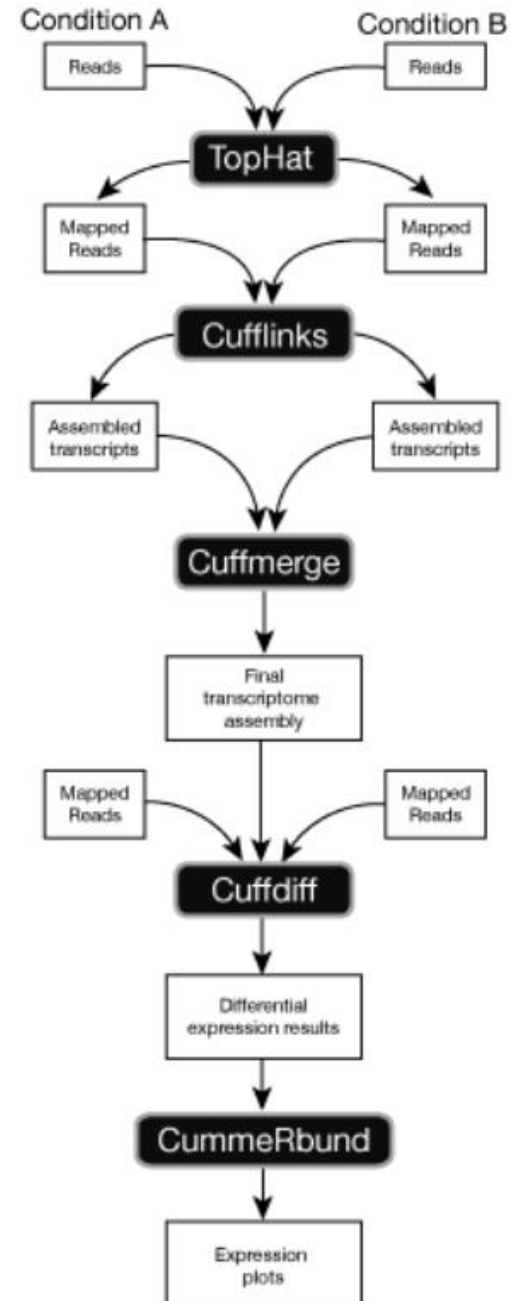
Table

	A	B	C	D
1	#gene	Dmel	Dsec	Dsim
2	128up	FBgn0010339	FBgn0176208	FBgn0182597
3	14-3-3epsilon	FBgn0020238	FBgn0170212	FBgn0186839
4	14-3-3zeta	FBgn0004907	FBgn0176067	FBgn0184006
5	140up	FBgn0010340	FBgn0180739	FBgn0191925
6	18w	FBgn0004364	FBgn0176867	FBgn0067532
7	26-29-p	FBgn0250848	FBgn0180299	FBgn0186151
8	312	FBgn0029514	FBgn0169330	FBgn0185324
9	4EHP	FBgn0053100	FBgn0178382	FBgn0189856
10	5-HT7	FBgn0004573	FBgn0167057	FBgn0188297
11	7B2	FBgn0041707	FBgn0165752	FBgn0191274
12	A16	FBgn0028965	FBgn0180777	FBgn0193504
13	ACC	FBgn0033246	FBgn0175616	FBgn0181977
14	AGO1	FBgn0262739	FBgn0178121	FBgn0084546
15	AIF	FBgn0031392	FBgn0178594	FBgn0194524
16	AP-1gamma	FBgn0030089	FBgn0168650	FBgn0188509
17	AP-1sigma	FBgn0030122	FBgn0181204	FBgn0192502

Analysis steps

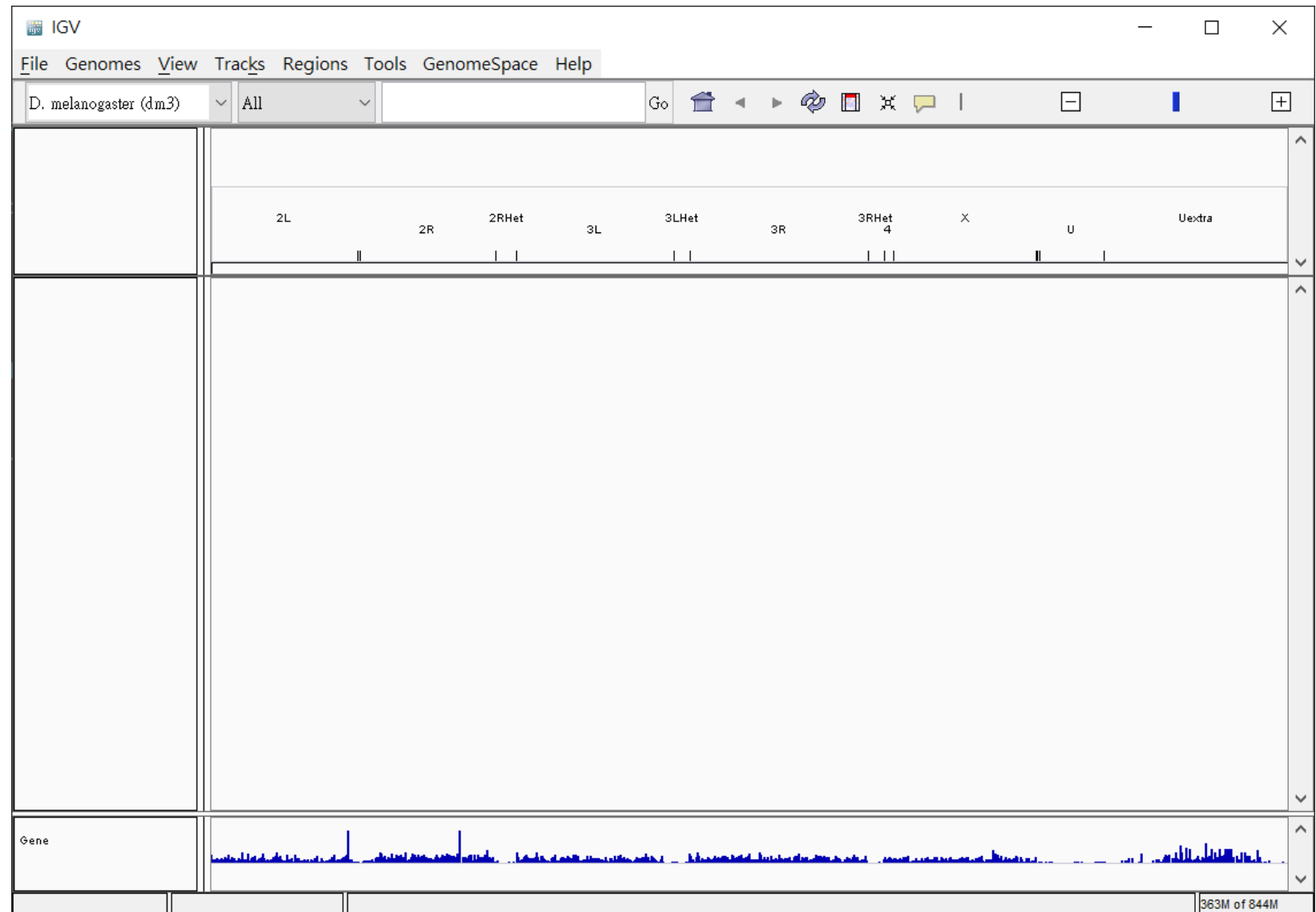


Analysis steps



IGV

- Tophat
- Samtools
- Cufflinks
- Cuffmerge
- Cuffdiff
- Use IGV
- CummeRbund



Reference genomes


- Reference genomes version
 - r5.48 (D. melanogaster) (Dm3)
 - r1.3 (D. simulans) (droSim1)
 - r1.3(D. sechellia) (droSec1)

Try

- Download data for many days

EMBL-EBI

ServicesResearchTrainingAbout us



Examples: [BN000065](#), [histone](#)
[Search](#)
[Advanced Sequence](#)

HomeSearch & BrowseSubmit & UpdateSoftwareAbout ENASupport

The new ENA Browser is now live, with improved features for searching & downloading data!
Please go to <https://www.ebi.ac.uk/ena/browser/view/SRR1973492> to see this record there.

Run: SRR1973492

[Contact Helpdesk](#)

Illumina HiSeq 2000 paired end sequencing; GSM1657333: Dsim male JP rep1; Drosophila simulans; RNA-Seq

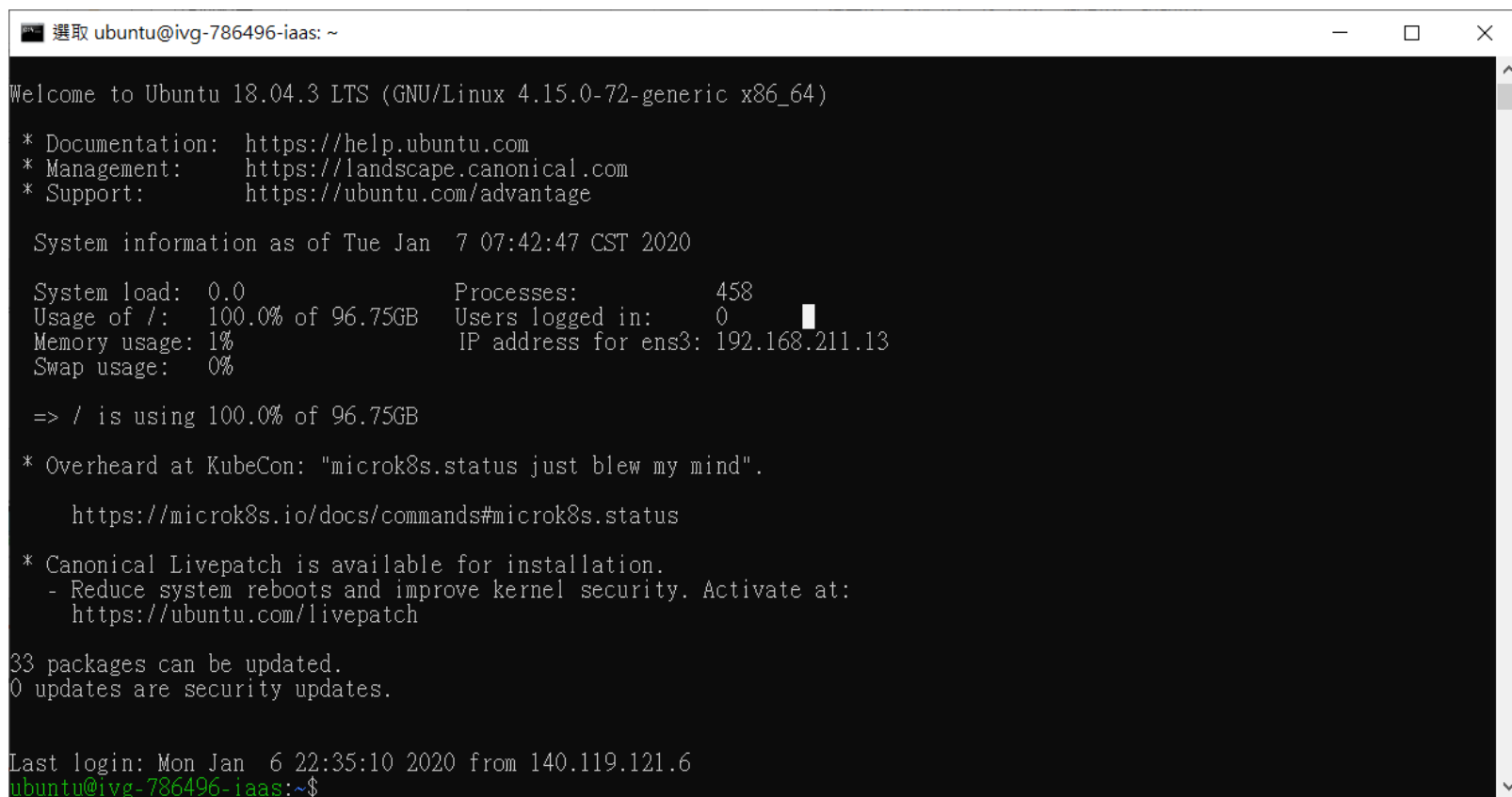
View: [XML](#)

Download: [XML](#)

Submitting Centre GEO	Platform ILLUMINA	Model Illumina HiSeq 2000	Read Count 28,225,417	Base Count 5,701,534,234
Library Layout PAIRED	Library Strategy RNA-Seq	Library Source TRANSCRIPTOMIC	Library Selection cDNA	Library Name
Broker Name NCBI				

Try

- **Ubuntu(R>3.6!!!!!!)**
- Oracle VM VirtualBox
- RMySQL
- Devtools
- Hmisc



```
ubuntu@ivg-786496-iaas: ~  
Welcome to Ubuntu 18.04.3 LTS (GNU/Linux 4.15.0-72-generic x86_64)  
  
* Documentation:  https://help.ubuntu.com  
* Management:    https://landscape.canonical.com  
* Support:        https://ubuntu.com/advantage  
  
System information as of Tue Jan  7 07:42:47 CST 2020  
  
System load:  0.0           Processes:            458  
Usage of /:   100.0% of 96.75GB Users logged in:     0  
Memory usage: 1%           IP address for ens3: 192.168.211.13  
Swap usage:   0%  
  
=> / is using 100.0% of 96.75GB  
  
* Overheard at KubeCon: "microk8s.status just blew my mind".  
  
    https://microk8s.io/docs/commands#microk8s.status  
  
* Canonical Livepatch is available for installation.  
  - Reduce system reboots and improve kernel security. Activate at:  
    https://ubuntu.com/livepatch  
  
33 packages can be updated.  
0 updates are security updates.  
  
Last login: Mon Jan  6 22:35:10 2020 from 140.119.121.6  
ubuntu@ivg-786496-iaas:~$
```


packages

- Cufflinks
- Cuffmerge
- Cummerbund
- Tophat
- Bowtie2 (NEED Build Index)
- samtools

```
ubuntu@ivg-786496-iaas: ~
System load: 0.01      Processes:      460
Usage of /: 100.0% of 96.75GB    Users logged in: 0
Memory usage: 1%      IP address for ens3: 192.168.211.13
Swap usage: 0%

=> / is using 100.0% of 96.75GB

* Overheard at KubeCon: "microk8s.status just blew my mind".
  https://microk8s.io/docs/commands#microk8s.status

* Canonical Livepatch is available for installation.
  - Reduce system reboots and improve kernel security. Activate at:
    https://ubuntu.com/livepatch

33 packages can be updated.
0 updates are security updates.

Last login: Tue Jan  7 07:42:47 2020 from 118.166.40.153
ubuntu@ivg-786496-iaas:~$ ls
R      SRR1952775_1.fastq      SRR1973487_1.fastq.gz  dm3_index.1.bt2      eg2.sam
SRR1952772_1.fastq      SRR1952775_2.fastq.gz  SRR1973487_2.fastq.gz  dm3_index.2.bt2      get-pip.py
SRR1952772_2.fastq      SRR1952776_1.fastq.gz  SRR1973488_1.fastq.gz  dm3_index.3.bt2      samtools-1.9
SRR1952772_tophat_out  SRR1952776_2.fastq.gz  SRR1973488_2.fastq.gz  dm3_index.4.bt2      samtools-1.9.tar.bz2
SRR1952773_1.fastq      SRR1952777_1.fastq.gz  SRR1973492_1.fastq.gz.st  dm3_index.rev.1.bt2  test_data
SRR1952773_2.fastq      SRR1952777_2.fastq.gz  bio                        dm3_index.rev.2.bt2  test_data.tar.gz
SRR1952774_1.fastq      SRR1973486_1.fastq.gz  bowtie2-2.3.5.1          droSec1.2b1t         tophat_out
SRR1952774_2.fastq      SRR1973486_2.fastq.gz  dm3.fasta                droSec1.fa           v2.3.5.1.tar.gz
ubuntu@ivg-786496-iaas:~$
```

Result

```
ubuntu@ivg-786496-iaas: ~  
19 root      20   0      0      0      0 S   0.0  0.0   0:00.00 cpuhp/2  
20 root      rt    0      0      0      0 S   0.0  0.0   0:00.16 watchdog/2  
21 root      rt    0      0      0      0 S   0.0  0.0   0:00.00 migration/2  
22 root      20   0      0      0      0 S   0.0  0.0   0:00.00 ksoftirqd/2  
24 root      0 -20     0      0      0 I   0.0  0.0   0:00.00 kworker/2:0H  
[71]+  Stopped                  top  
ubuntu@ivg-786496-iaas:~$ tophat -I 1000 -i 20 -o SRR1952772_tophat_out dm3_index SRR1952772_1.fastq SRR1952772_2.fastq  
[2020-01-06 21:30:59] Beginning TopHat run (v2.1.1)  
-----  
[2020-01-06 21:30:59] Checking for Bowtie  
Bowtie version:      2.3.4.1  
[2020-01-06 21:30:59] Checking for Bowtie index files (genome)..  
[2020-01-06 21:30:59] Checking for reference FASTA file  
Warning: Could not find FASTA file dm3_index.fa  
[2020-01-06 21:30:59] Reconstituting reference FASTA file from Bowtie index  
Executing: /usr/bin/bowtie2-inspect dm3_index > SRR1952772_tophat_out/tmp/dm3_index.fa  
[2020-01-06 21:31:05] Generating SAM header for dm3_index  
[2020-01-06 21:31:05] Preparing reads  
[FAILED]  
Error retrieving prep_reads info.  
ubuntu@ivg-786496-iaas:~$ tophat -I 1000 -i 20 -o SRR1952772_tophat_out dm3_index SRR1952772_1.fastq SRR1952772_2.fastq  
[2020-01-06 21:49:49] Beginning TopHat run (v2.1.1)  
-----  
[2020-01-06 21:49:49] Checking for Bowtie  
Bowtie version:      2.3.4.1  
[2020-01-06 21:49:49] Checking for Bowtie index files (genome)..  
[2020-01-06 21:49:49] Checking for reference FASTA file  
Warning: Could not find FASTA file dm3_index.fa
```

(X____X)