

Genomes and Unix

“Introductory advanced course”

Václav Janoušek, Libor Mořkovský

Genomes

What is a genome?

WAR
And
PEACE



LEO TOLSTOY

A NEW TRANSLATION BY
RICHARD PEVEAR AND LARissa VOLOKHOVSKY

WAR
And
PEACE

LEO
TOLSTOY



RICHARD PEVEAR
AND LARissa VOLOKHOVSKY





What is a genome?

- sequence
- physical map
- annotations
- versioned reference

impossible en fait ou en droit. S'il s'agissait d'un enfant rceau, d'un absent ou d'une femme, il était nécessaire que quelqu'un prit en main la cause de ce plaideur, qui ne pouvait pas présenter lui-même sa défense. Si l'Etat était également intéressé à une affaire, si par exemple il s'agissait d'un procès de haute trahison, il ne fallait pas qu'en un seul et aussi grave la défense de l'intérêt public fût abandonnée à l'initiative privée. La république désignait donc des orateurs chargés de la représenter, et de soutenir l'accusation. Dans ces cas, il avait bien fallu déroger à la règle, et admis

What is a genome?

- sequence
- physical map
- annotations
- versioned reference

CONTENTS

BOOK I.	
CHAPTER	PAGE
I. Introduction	1
II. On War and Right	17
III. Inquiry into the Lawfulness of War	31
IV. The Division of War into Public and Private, and the Nature of Sovereign Power	55
BOOK II.	
I. Defense of Person and Property	73
II. The General Rights of Things	85
III. On the Original Acquisition of Things, and the Right of Property in Seas and Rivers	103
IV. Title to Desert Lands by Occupancy, Possession, and Prescription	109
IX. In What Cases Jurisdiction and Property Cease	117
X. The Obligation Arising from Property	125
XI. On Promises	131
XII. On Contracts	144
XIII. On Oaths	160
XV. On Treaties and on Engagements Made by Delegates Exceeding Their Powers	166
XVI. The Interpretation of Treaties	175
XVII. On Duties Contracted by Injury, and the Obligation to Repair Them	195
XVIII. On the Rights of Embassies	208
XIX. On the Right of Burial	213
XX. On Punishments	220
XXI. On the Communication of Punishment	226
XXII. On the Unjust Causes of War	237
XXIII. On Doubtful Causes	274
XXIV. Precautions against Rashly Engaging in War, Even upon Just Grounds	280
BOOK III.	
I. What Is Lawful in War	309
II. In What Manner the Law of Nations Renders the Property of Subjects Answerable for the Debts of Sovereigns. The Nature of Reprisals	307
III. On Just or Solemn War According to the Law of Nations on Declarations of War	314 (xlii)

What is a genome?

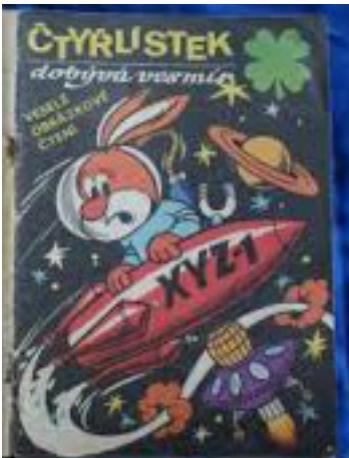
- sequence
- physical map
- annotations
- versioned reference

About eight years ago Mr Starling wrote about Mr Brown who finds tiny body parts which at first might be human body parts to us readers on the frosty pavement outside the butcher's shop, but which turn out to be made of creamy-white ivory and which he puts in an empty plastic butcher's bag anyway and keeps in his fridge until the evening (which) he spends piecing and gluing the parts back together and thinking about Mr Chalmers a museum collector who as a boy had carved a small flat relief of the head of the Roman Emperor Mr Caesar and about Mr Fricke who turned raw steel into gilt-framed masterpieces in the nineteenth century and about the thieves who had burgled the museum in the first place earlier that day and dropped the broken limbs as they escaped and imagining Mr Chalmers flattened but unharmed in a childhood accident like his relief of Mr Caesar, and who in the morning begins to write the story I read as a child called *Flat Stanley* by Jeff Brown which begins "Breakfast was ready"

who would fit its bones tightly & with inflex

What is a genome?

- sequence
- physical map
- annotations
- versioned reference



Where to find genomes?

- Ensembl - download - FTP

Show	10 ▾ entries	Show/hide columns						Filter	
★	Species	DNA (FASTA)	cDNA (FASTA)	ncRNA (FASTA)	Protein sequence (FASTA)	Whole databases	Variation (VCF)	Regulation (GFF)	BAM
Y	Human <i>Homo sapiens</i>	FASTA	FASTA	FASTA	FASTA	MySQL	VCF	Regulation (GFF)	BAM
Y	Mouse <i>Mus musculus</i>	FASTA	FASTA	FASTA	FASTA	MySQL	VCF	Regulation (GFF)	BAM
Y	Zebrafish <i>Danio rerio</i>	FASTA	FASTA	FASTA	FASTA	MySQL	VCF	-	BAM
	Alpaca <i>Vicugna pacos</i>	FASTA	FASTA	FASTA	FASTA	MySQL	-	-	
	Anole lizard <i>Anolis carolinensis</i>	FASTA	FASTA	FASTA	FASTA	MySQL	-	-	BAM
	Armadillo <i>Dasyurus novemcinctus</i>	FASTA	FASTA	FASTA	FASTA	MySQL	-	-	BAM
	Bushbaby <i>Otolemur garnettii</i>	FASTA	FASTA	FASTA	FASTA	MySQL	-	-	
	C.intestinalis <i>Ciona intestinalis</i>	FASTA	FASTA	FASTA	FASTA	MySQL	-	-	
	C.savignyi <i>Ciona savignyi</i>	FASTA	FASTA	FASTA	FASTA	MySQL	-	-	
	Caenorhabditis elegans <i>Caenorhabditis elegans</i>	FASTA	FASTA	FASTA	FASTA	MySQL	-	-	

Where to find genomes?

- UCSC - Genomes - Downloads



Jim Kent

How genomes are created?

été impossible en fait ou en droit. S'il s'

quelqu'un prit en main la cause de ce j
été impossible en fait ou en droit. S'il s'
trait rebelle à la parole, il devenait bie
D'ailleurs il y avait des cas où l'o
gés de la représenter, et de soutenir l'
s'en occuper. directement intéressé à une affaire,
tre que le plaideur se pouvait faire repi
ne devait pas présenter lui-même sa dé
é devait pas présenter lui-même sa défe

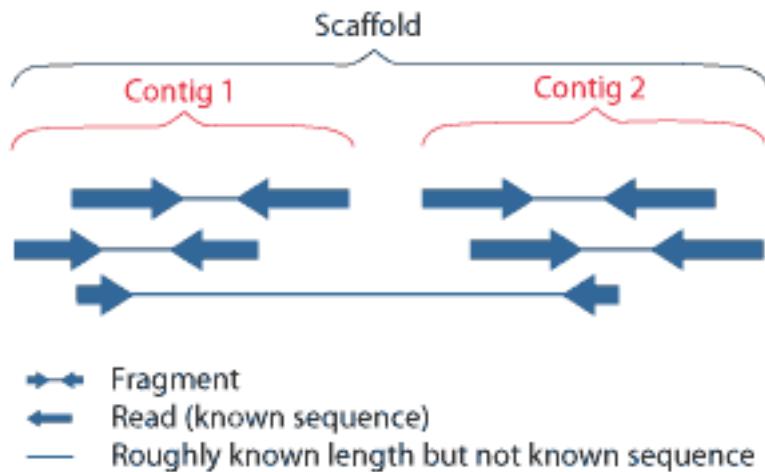
été impossible en fait ou en droit. S'il s'
fait ou en droit. S'il s'agissait d'un enf~~e~~
d'un enfant au

été impossible en fait ou en droit. S'il s'

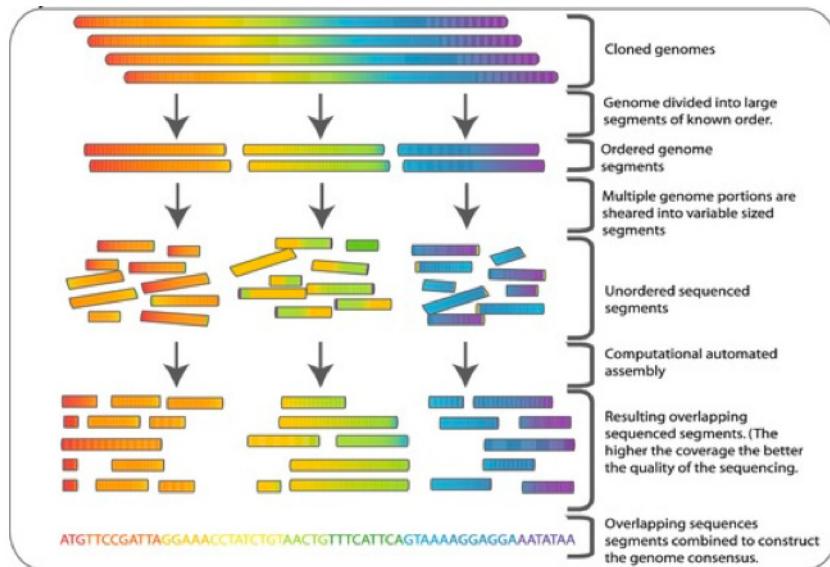
fait ou en droit. S'il s'agissait d'un enfant au

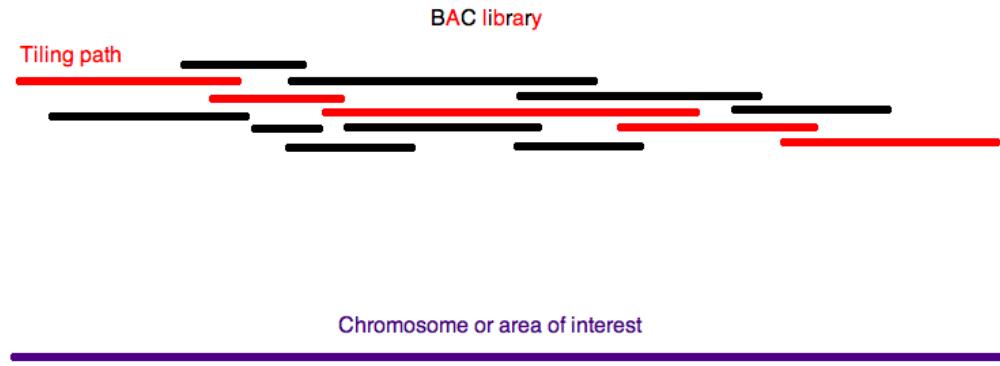
été impossible en fait ou en droit. S'il s'agissait d'un enfant au

contig

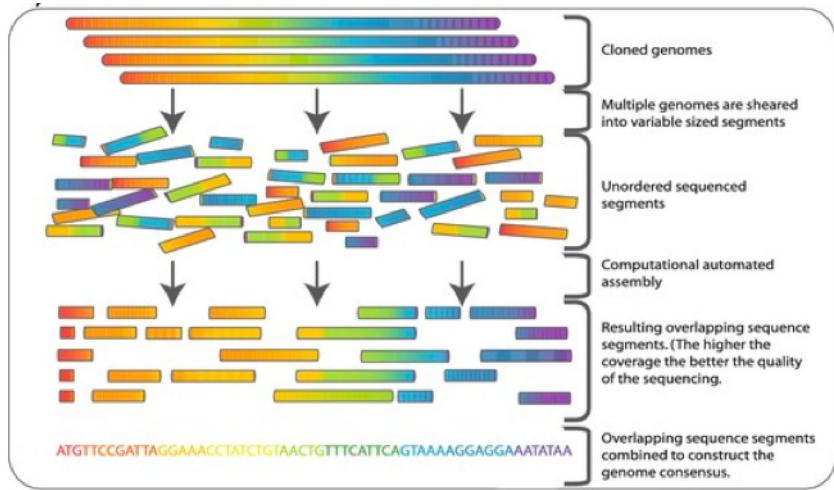


The old way:





The new way:



This is easy!
Can we have a genome of *whatever*?

Illumina sales agent says: “*Of course, in four days.*”

Illumina sales agent says: "Of course, in four days."



Sequence: assembly quality

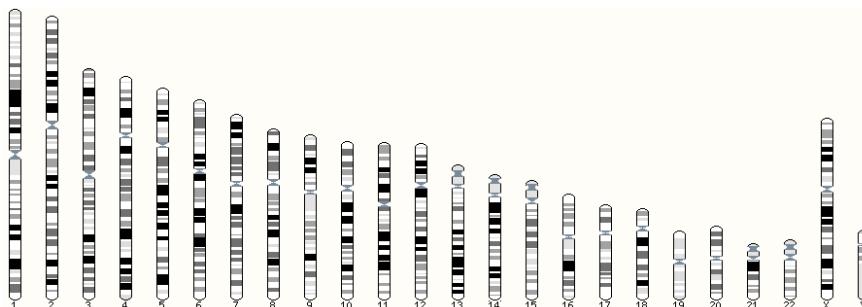
- coverage
- longest contig size
- N50 or N90

tous ces cas, il avait bien fallu déroger à la règle, et c'était contre que le plaideur se pouvait faire représenter par autre.

De même, s'il s'agissait d'un laboureur, d'un vigneron, d'un matelot ou d'un soldat, dont la langue indocile se refusait rebelle à la parole, il devenait bien difficile d'approuver la loi. C'eût été vraiment perdre le temps des juges, et puis se moquer d'eux que de leur produire un plaideur stupide et incapable d'exposer le premier mot de son affaire. N'allons pas d'ailleurs une injustice criante que de pourrir des soldats blanchis sous le harnais, qui dans maints combats s'étaient couverts d'une glorieuse sueur, fussions au contraire sans défense aux attaques d'un vil sycophante, et de l'autre

Physical map: what is the “new” way?

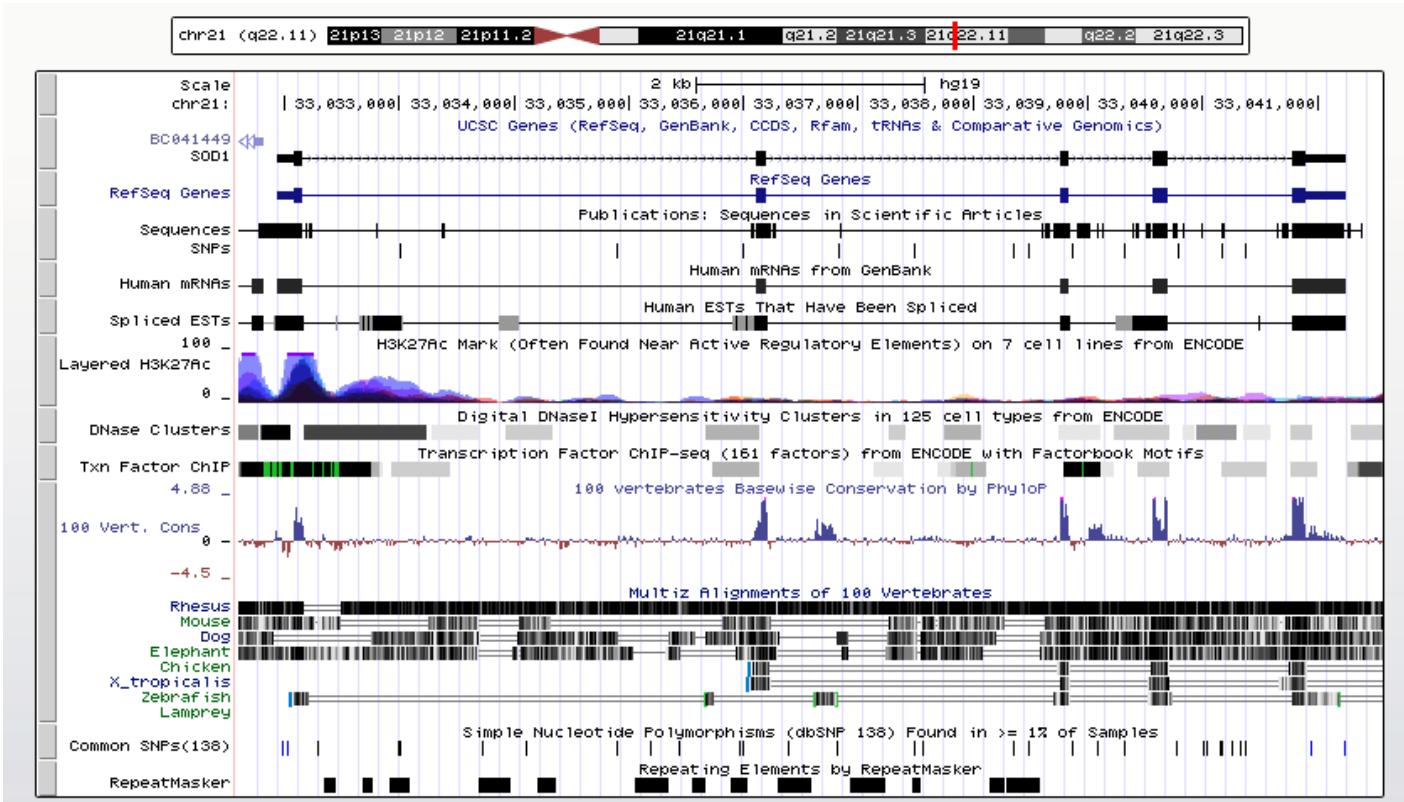
- breeding, breeding, breeding
- RAD tags do help



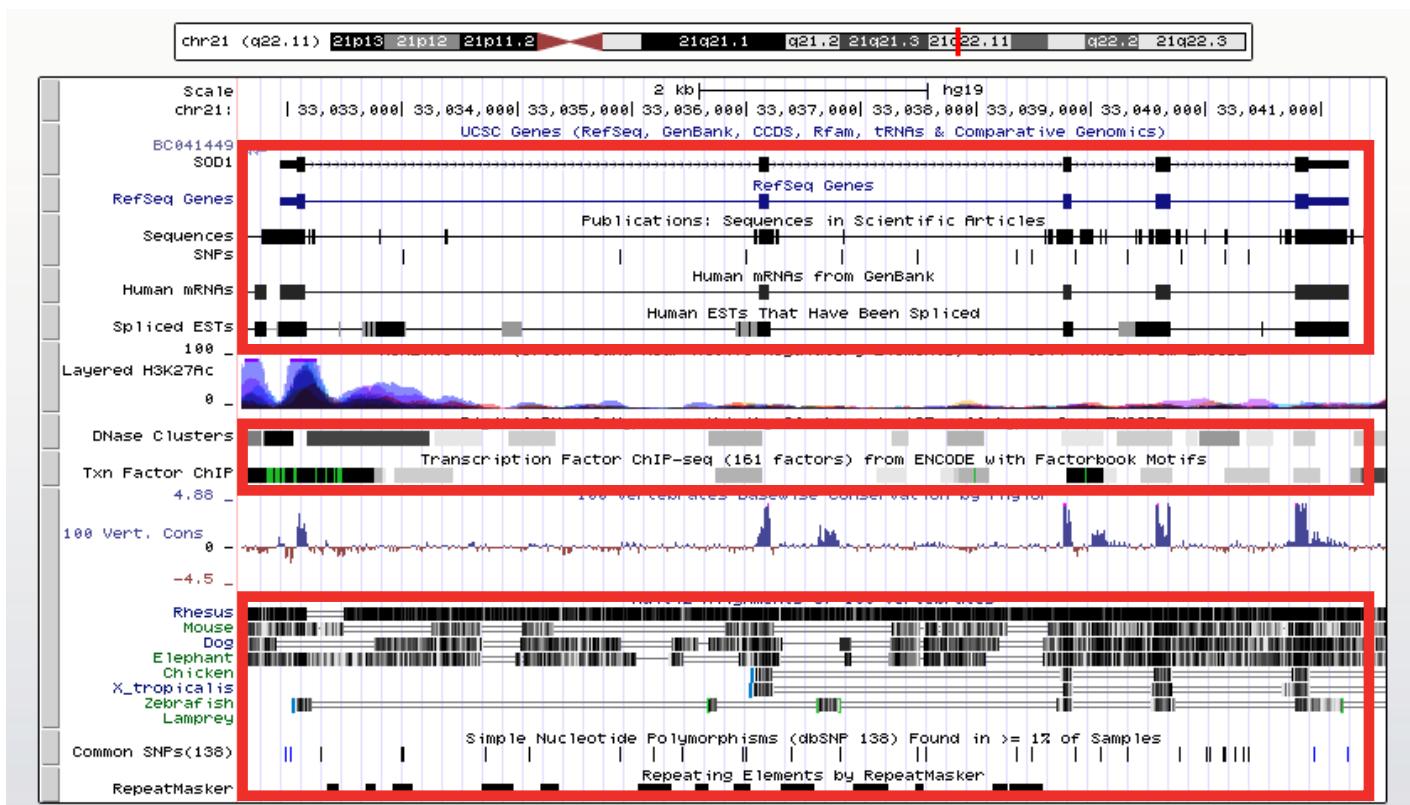
CONTENTS

BOOK I.	
CHAPTER	PAGE
I. Introduction	1
II. On War and Right	17
III. Inquiry into the Lawfulness of War	31
III. The Division of War into Public and Private, and the Nature of Sovereign Power	55
BOOK II.	
CHAPTER	PAGE
I. Defense of Persons and Property	73
II. The General Rights of Things	75
III. On the Original Acquisition of Things, and the Right of Property in Seas and Rivers	83
IV. Title to Desert Lands by Occupancy, Possession, and Prescription	109
X. In What Cases Jurisdiction and Property Cease	117
XI. On Obligations Arising from Property	123
XII. On Contracts	141
XIII. On Oaths	144
XV. On Treaties and on Engagements Made by Delegates Exceeding Their Powers	166
XVI. The Remedies of Treaties	176
XVII. On Damages Occasioned by Injury, and the Obligation to Repay Them	195
XVIII. On the Rights of Embassies	203
XIX. On the Right of Burial	213
XX.	219
XXI. On the Communication of Punishment	226
XXII. On the Unjust Causes of War	237
XXIII. On Doubtful Causes	274
XXIV. Precautions against Rashly Engaging in War, Even upon Just Grounds	280
BOOK III.	
I. What Is Lawful in War	392
II. In What Manner the Law of Nations Renders the Property of Subjects Answerable for the Debts of Sovereigns. The Nature of Retaliations	397
III. On Just or Solemn War According to the Law of Nations on Declarations of War	394
(viii)	395

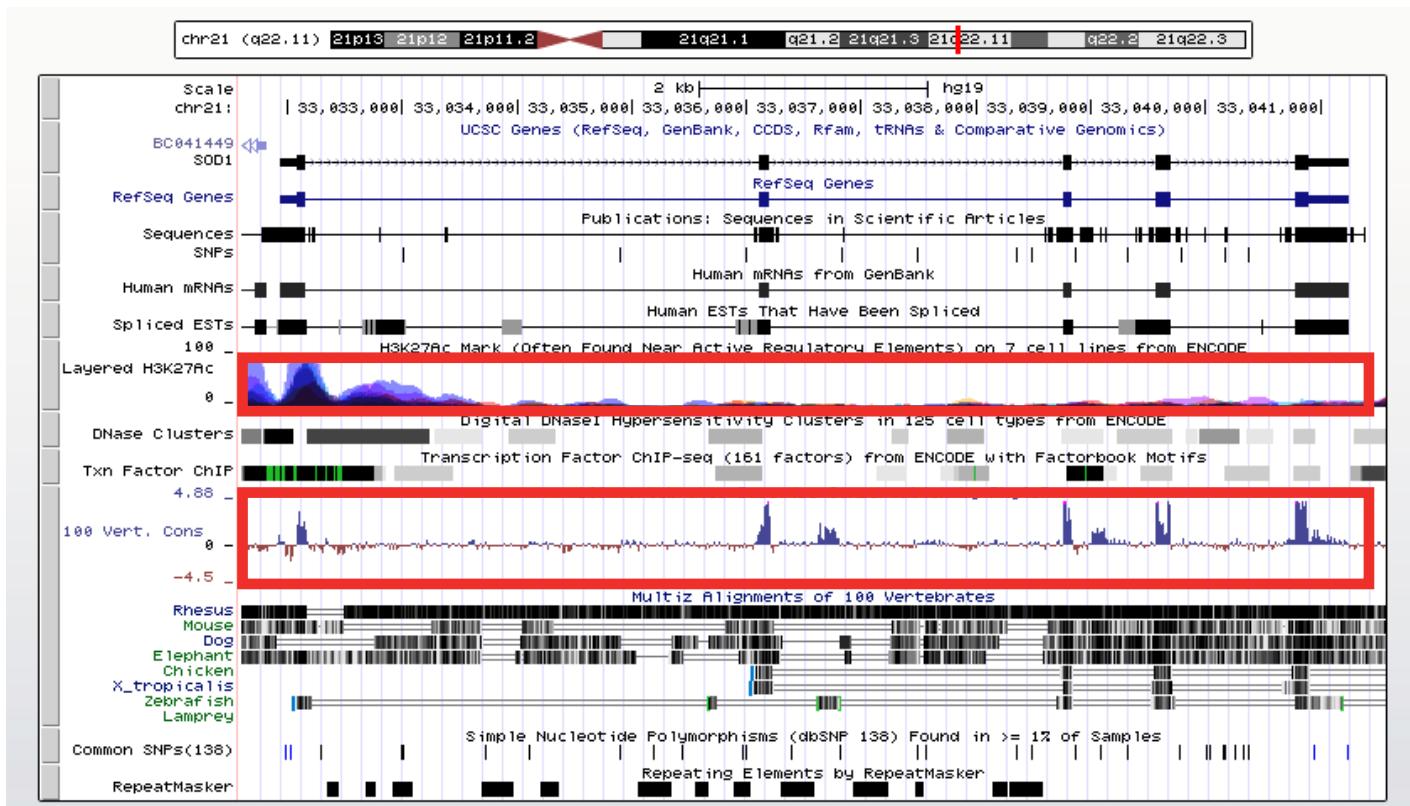
Annotations:



Annotations: "features" (bed, gff)



Annotations: “continuous valued” (wig)



Annotations: how do we get them?

- sequence similarity: ESTs, RNASeq
- sequence similarity: other species
- mathemagics: prediction models

Annotations: how do we get them?

- sequence similarity: ESTs, RNASeq
- sequence similarity: other species
- mathemagics: prediction models

Annotations: how do we get them?

- sequence similarity: ESTs, RNASeq
- sequence similarity: other species
- mathemagics: prediction models

Formats: FASTA

>chr1

GGAGAACGAGAACACGACATCTCcCTCTTCgTCAaGGCTGGCaGtGaTGGGgAAaGCATTGGGAAC
TGCCCcgTTCTCTCAGCGCCTTTATGATCCTGTGGCTAAAGGTGTCATATTAATGTCACAACCg
TGGATC

>chr2

GGAGGCTCACGCTGATATTATTGAAAGGATTGtCTCCTGGGAGACTGGGGTGAActtGTTGGTGG
TAGCAAGAGAAGCGATAGAATGATTTTTGGTTGGCTTTTTttCCTTCTACCTTCCCCTGCT
TGTTGAATAGCTGGTGTAGAAAAGGCCCTTGAGAATACGTGTTCAAGCTTTGGTTGTTCTGGG
GATACTGAC

Formats: FASTQ

```
@G59B7NP01B3ox5
AAGCCTCACCTACACTTAAGAGCTGAGTAGAATAAGCAAATGCTAATGCACCAGAAAGAGG...
+
HGGGIIGG999IIHGGGHIIIIIIIIIIIIIIIIIIIEEEIDD99888=A<::=B?==:466...
@G59B7NP01CPQYF
CCATGAGTGGATAGCAGGTTATGGACACAGAGGTTAGAGGAAGCAAGGAAGGAGAAGTAATGAG...
+
IIIIIIIIIIIIICCCIIIIIIIG; ; ; FC@8877BC//. . . --. 58?C@<?EEII...
```

Formats: BED

chrom	start	end	name	score	strand
chr7	127471196	127472363	Pos1	0	+
chr7	127472363	127473530	Pos2	0	+
chr7	127473530	127474697	Pos3	0	+
chr7	127474697	127475864	Pos4	0	+
chr7	127475864	127477031	Neg1	0	-
chr7	127477031	127478198	Neg2	0	-
chr7	127478198	127479365	Neg3	0	-
chr7	127479365	127480532	Pos5	0	+
chr7	127480532	127481699	Neg4	0	-

GFF and VCF are just more complex cousins of BED

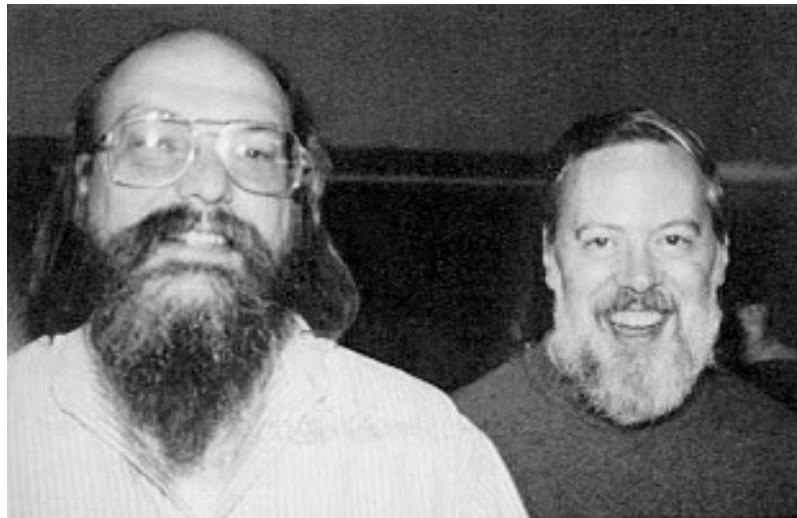
Unix

“Those who don’t understand Unix are condemned to reinvent it, poorly.”

Henry Spencer

“Unix never says ‘please’.”

Rob Pike



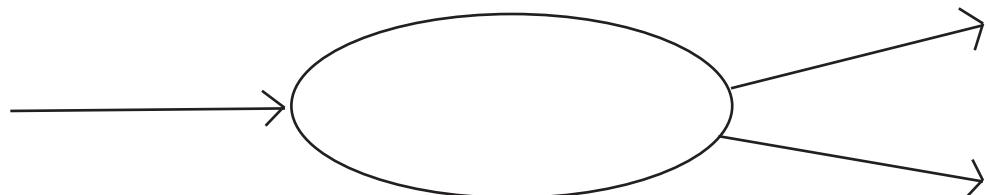
Ken Thompson and Dennis Ritchie

What is it all about?

- flexibility
- conciseness
- automation

Flexibility: **every** program has

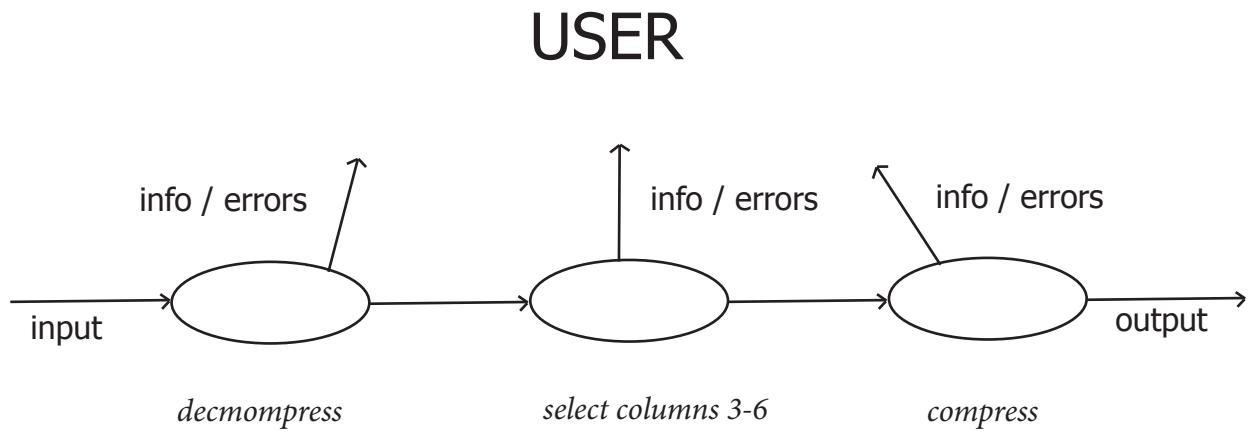
- standard input
- standard output
- standard error



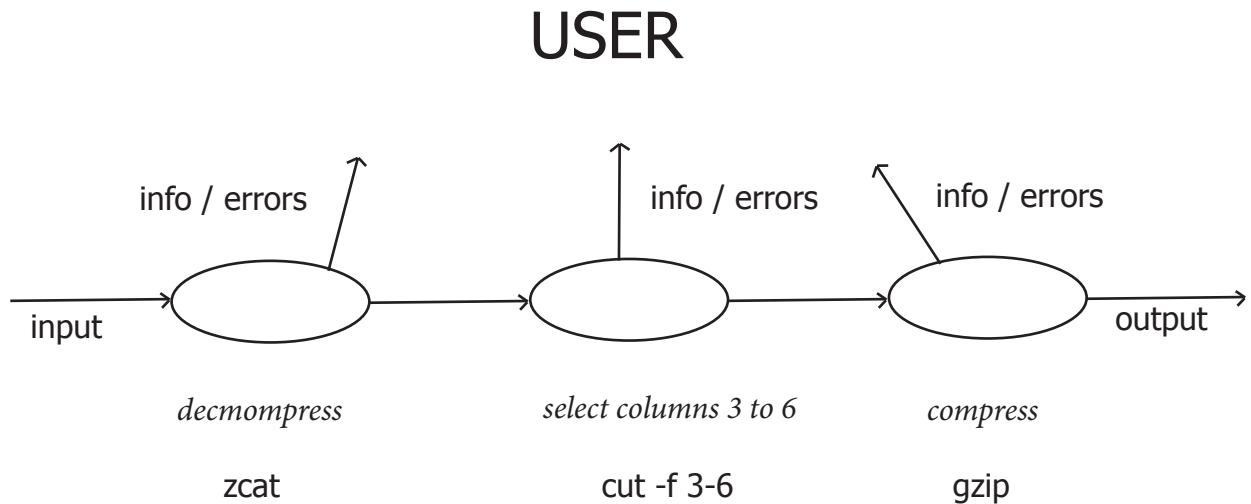
Flexibility: **every** “unix” program

- reads data line by line
- outputs data line by line
- does one simple operation

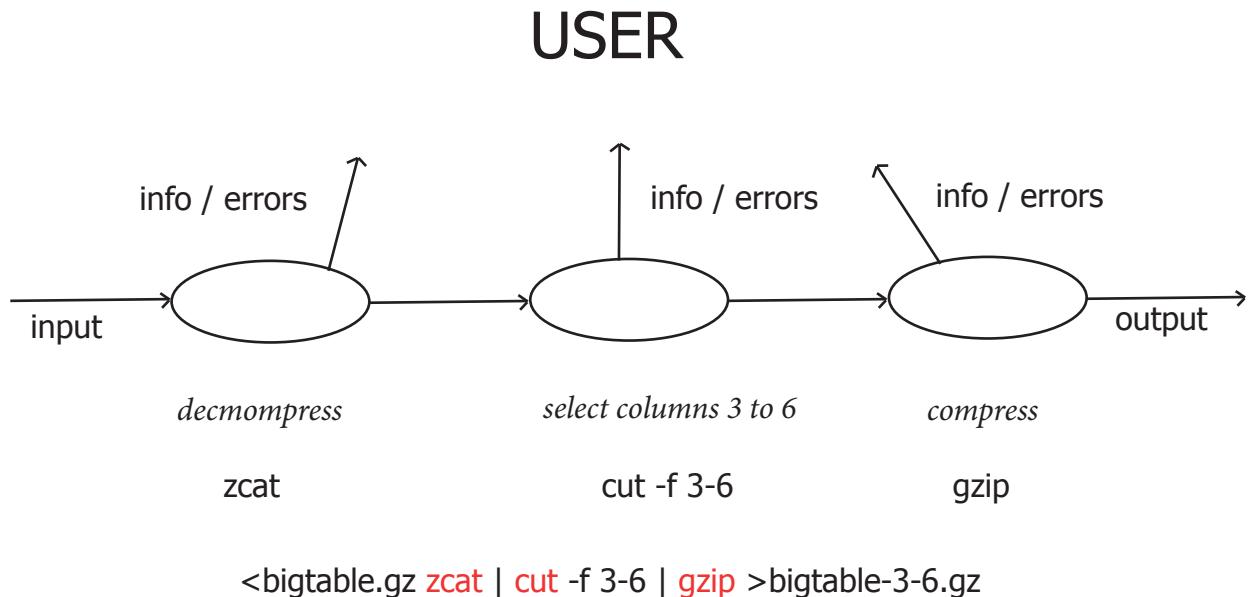
Flexibility: programs can be chained



Conciseness: “with just 30 characters you can do almost anything”



Automation: every operation can be stored in a text file



And what is Linux then?

