

# Genomes and Unix

*“Introductory advanced course”*

Václav Janoušek, Libor Mořkovský

# Genomes

What is a genome?

# WAR *And* PEACE



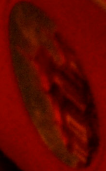
LEO TOLSTOY

A NEW TRANSLATION BY

RICHARD PEVELAR AND LARISA TOLSTOVSKY

WAR  
*And*  
PEACE

LEO  
TOLSTOY



RICHARD PEVELAR  
AND LARISA TOLSTOVSKY





# What is a genome?

- sequence
- physical map
- annotations
- versioned reference

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



# What is a genome?

- sequence
- physical map
- annotations
- versioned reference

CONTENTS	
BOOK I.	
CHAPTER	PAGE
Introduction . . . . .	1
I. On War and Right . . . . .	17
II. 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.	
I. Detest 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 . . . . .	123
XI. On Promises . . . . .	131
XII. On Contracts . . . . .	144
XIII. On Oaths . . . . .	150
XV. On Treaties and on Engagements Made by Delegates Exceeding Their Powers . . . . .	166
XVI. The Interpretation of Treaties . . . . .	175
XVII. On Damages Constrained 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 . . . . .	267
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 . . . . .	290
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
(xiii)	

# What is a genome?

- sequence
- physical map
- annotations
- versioned reference

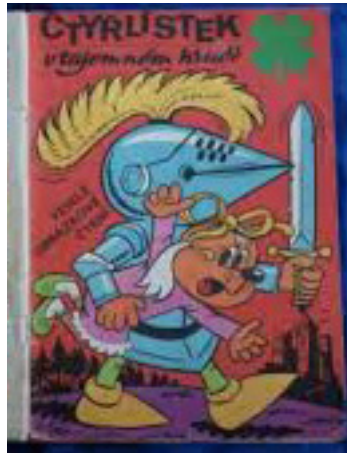
About eight years ago Mr Starling wrote about Mr Brown who finds tiny body parts which at first might be human body parts to us readers on the frosty pavement outside the butcher's shop, but which turn out to be made of creamy-white ivory and which he puts in an empty plastic butcher's bag anyway and 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 Frick 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.

which begins "Breakfast was ready" by Jeff Brown  
who would hit into James flatly & water his glass



# What is a genome?

- sequence
- physical map
- annotations
- versioned reference



# Where to find genomes?

- Ensembl - download - FTP

Show	▼ entries	Show/hide columns						Filter	
★	Species	DNA (FASTA)	cDNA (FASTA)	ncRNA (FASTA)	Protein sequence (FASTA)	Whole databases	Variation (VCF)	Regulation (GFF)	BAM
Y	<b>Human</b> <i>Homo sapiens</i>	<a href="#">FASTA</a>	<a href="#">FASTA</a>	<a href="#">FASTA</a>	<a href="#">FASTA</a>	<a href="#">MySQL</a>	<a href="#">VCF</a>	<a href="#">Regulation (GFF)</a>	<a href="#">BAM</a>
Y	<b>Mouse</b> <i>Mus musculus</i>	<a href="#">FASTA</a>	<a href="#">FASTA</a>	<a href="#">FASTA</a>	<a href="#">FASTA</a>	<a href="#">MySQL</a>	<a href="#">VCF</a>	<a href="#">Regulation (GFF)</a>	<a href="#">BAM</a>
Y	<b>Zebrafish</b> <i>Danio rerio</i>	<a href="#">FASTA</a>	<a href="#">FASTA</a>	<a href="#">FASTA</a>	<a href="#">FASTA</a>	<a href="#">MySQL</a>	<a href="#">VCF</a>	-	<a href="#">BAM</a>
	<b>Alpaca</b> <i>Vicugna pacos</i>	<a href="#">FASTA</a>	<a href="#">FASTA</a>	<a href="#">FASTA</a>	<a href="#">FASTA</a>	<a href="#">MySQL</a>	-	-	-
	<b>Anole lizard</b> <i>Anolis carolinensis</i>	<a href="#">FASTA</a>	<a href="#">FASTA</a>	<a href="#">FASTA</a>	<a href="#">FASTA</a>	<a href="#">MySQL</a>	-	-	<a href="#">BAM</a>
	<b>Armadillo</b> <i>Dasypus novemcinctus</i>	<a href="#">FASTA</a>	<a href="#">FASTA</a>	<a href="#">FASTA</a>	<a href="#">FASTA</a>	<a href="#">MySQL</a>	-	-	<a href="#">BAM</a>
	<b>Bushbaby</b> <i>Otlemur garnettii</i>	<a href="#">FASTA</a>	<a href="#">FASTA</a>	<a href="#">FASTA</a>	<a href="#">FASTA</a>	<a href="#">MySQL</a>	-	-	-
	<b>C.intestinalis</b> <i>Ciona intestinalis</i>	<a href="#">FASTA</a>	<a href="#">FASTA</a>	<a href="#">FASTA</a>	<a href="#">FASTA</a>	<a href="#">MySQL</a>	-	-	-
	<b>C.savignyi</b> <i>Ciona savignyi</i>	<a href="#">FASTA</a>	<a href="#">FASTA</a>	<a href="#">FASTA</a>	<a href="#">FASTA</a>	<a href="#">MySQL</a>	-	-	-
	<b>Caenorhabditis elegans</b> <i>Caenorhabditis elegans</i>	<a href="#">FASTA</a>	<a href="#">FASTA</a>	<a href="#">FASTA</a>	<a href="#">FASTA</a>	<a href="#">MySQL</a>	-	-	-
Showing 1 to 10 of 66 entries									

# 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

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,

ne devait pas présenter lui-même sa dé  
tre que le plaideur se pouvait faire rep

ne devait pas présenter lui-même sa défe



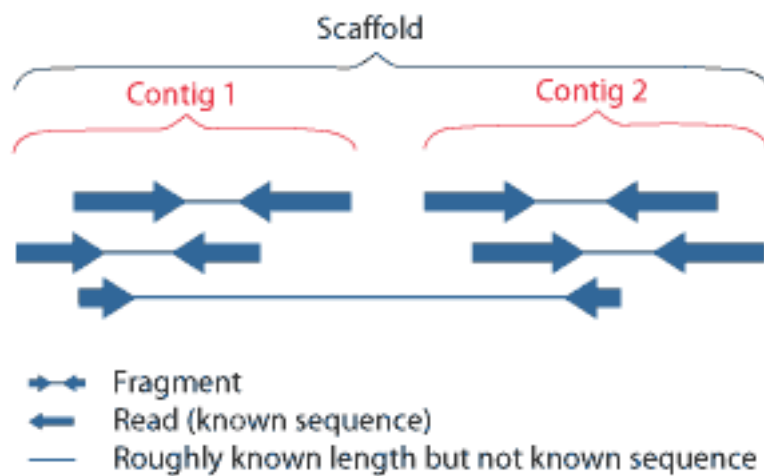
été impossible en fait ou en droit. S'il s'  
fait ou en droit. S'il s'agissait d'un enf  
d'un enfant au

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

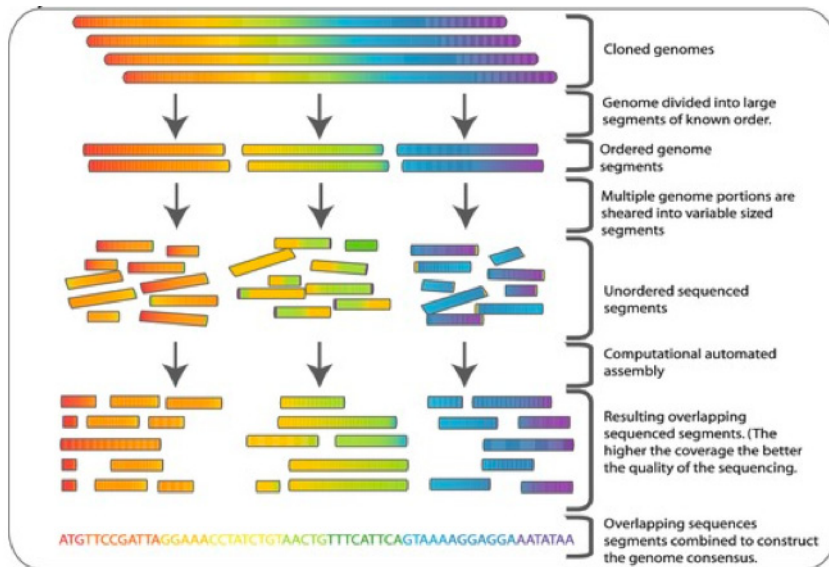
- fait ou en droit. S'il s'agissait d'un enf
- d'un enfant au
- 
- 
- 
- 
- 
- 

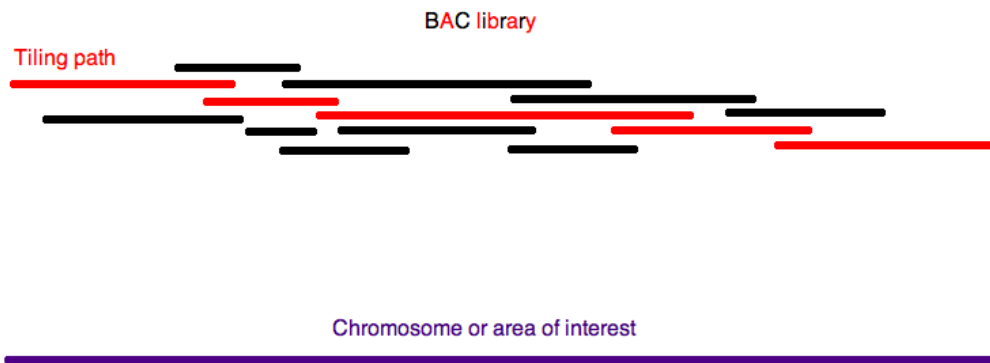
été impossible en fait ou en droit. S'il s'agissait d'un enfant au

**contig**

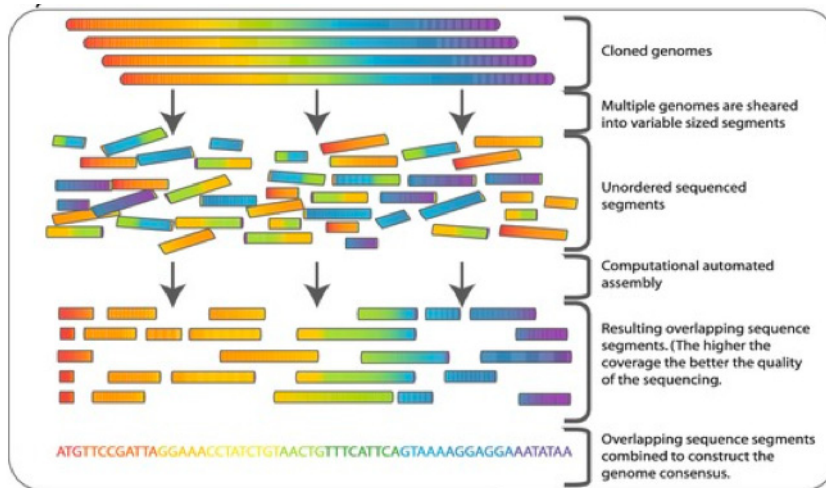


# The old way:





# 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."*

大スクリーンで蘇る文芸映画の超大作！

◎東京国際映画祭 第15回記念大会 特別上映作品

# 戦争と平和

全長7時間を超える4部作完全版 ただ1日限りの一挙上映

第1部「アンドレイ・ボルクンスキー」  
第2部「ナターシャ・ロストワ」  
第3部「1812年」  
第4部「ヒエール・ベズーホフ」

◎2002年11月3日 日曜日 ◎12:00開場／12:30開映  
◎東京・渋谷 Bunkamuraオーチャードホール

第4回モスクワ国際映画祭グランプリ・審査員特別賞・1967年度米アカデミー外国語映画賞  
原作 レフ・トルストイ 製作・監督 脚本 主演・ナレーター セルゲイ・ボンダルチュク 共演脚本 ワシリー・ソロヴィエフ  
撮影 アナトーリ・ペトリツキー 美術 ミハイル・ボグダノフ 音楽 アレクサンドル・ミハエエスコ 音楽 ウィエチスラフ・オプテンニコフ  
出演 リュドミラ・サヴェーリェワ・ウィヤチスラフ・チーホノフ・アナスタシア・ヴェルシンスカヤ  
1965～1967年 カラー シネスコ モノラル ロシア語・日本語字幕（英語字幕はついていません） 約7時間50分

# Sequence: assembly quality

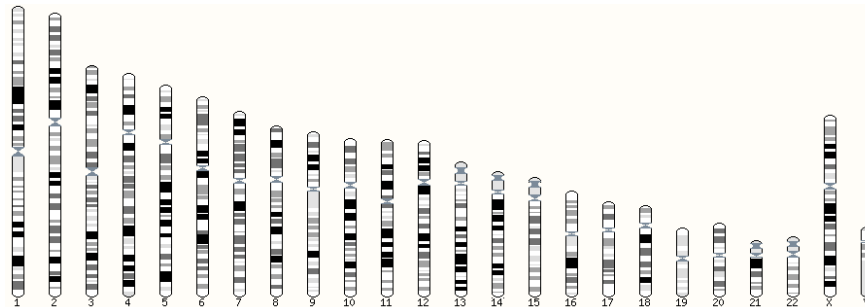
- coverage
- longest contig size
- N50 or N90

tous ces cas, il avait bien fallu déroger à la règle, et admettre que le plaideur se pouvait faire représenter par autrui.

De même, s'il s'agissait d'un laboureur, d'un vigneron, d'un matelot ou d'un soldat, dont la langue indocile se montrait rebelle à la parole, il devenait bien difficile d'appliquer la loi. C'eût été vraiment perdre le temps des juges, et se moquer d'eux que de leur produire un plaideur absolument incapable d'exposer le premier mot de son affaire. N'était-ce pas d'ailleurs une injustice criante que de priver ces soldats blanchis sous le harnais, qui dans maints combats s'étaient couverts d'une glorieuse sueur, fussent exposés sans défense aux attaques d'un vil sycophante, et se vissent

# Physical map: what is the “new” way?

- breeding, breeding, breeding
- RAD tags do help



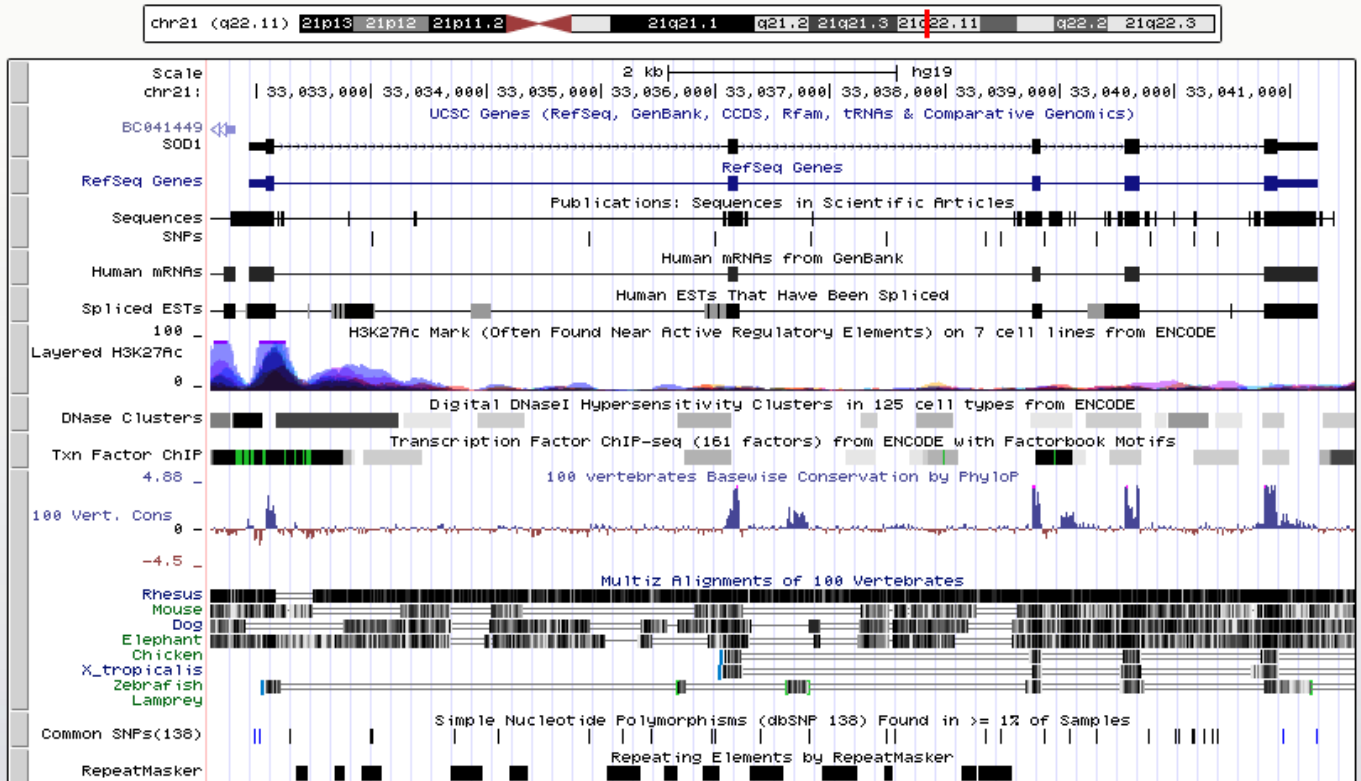
## CONTENTS

BOOK I.		Page
CHAPTER		
I.	On War and Right . . . . .	17
II.	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.		Page
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 . . . . .	123
XI.	On Promises . . . . .	131
XII.	On Contracts . . . . .	144
XIII.	On Oaths . . . . .	150
XV.	On Treaties and on Engagements Made by Delegates Exceeding Their Powers . . . . .	166
XVI.	The Interpretation of Treaties . . . . .	175
XVII.	On Damages Occasioned by Injury, and the Obligation to Repair Them . . . . .	195
XVIII.	On the Rights of Embassies . . . . .	209
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 . . . . .	267
XXIII.	On Doubtful Causes . . . . .	274
XXIV.	Precautions against Rashly Engaging in War, Even upon Just Grounds . . . . .	280

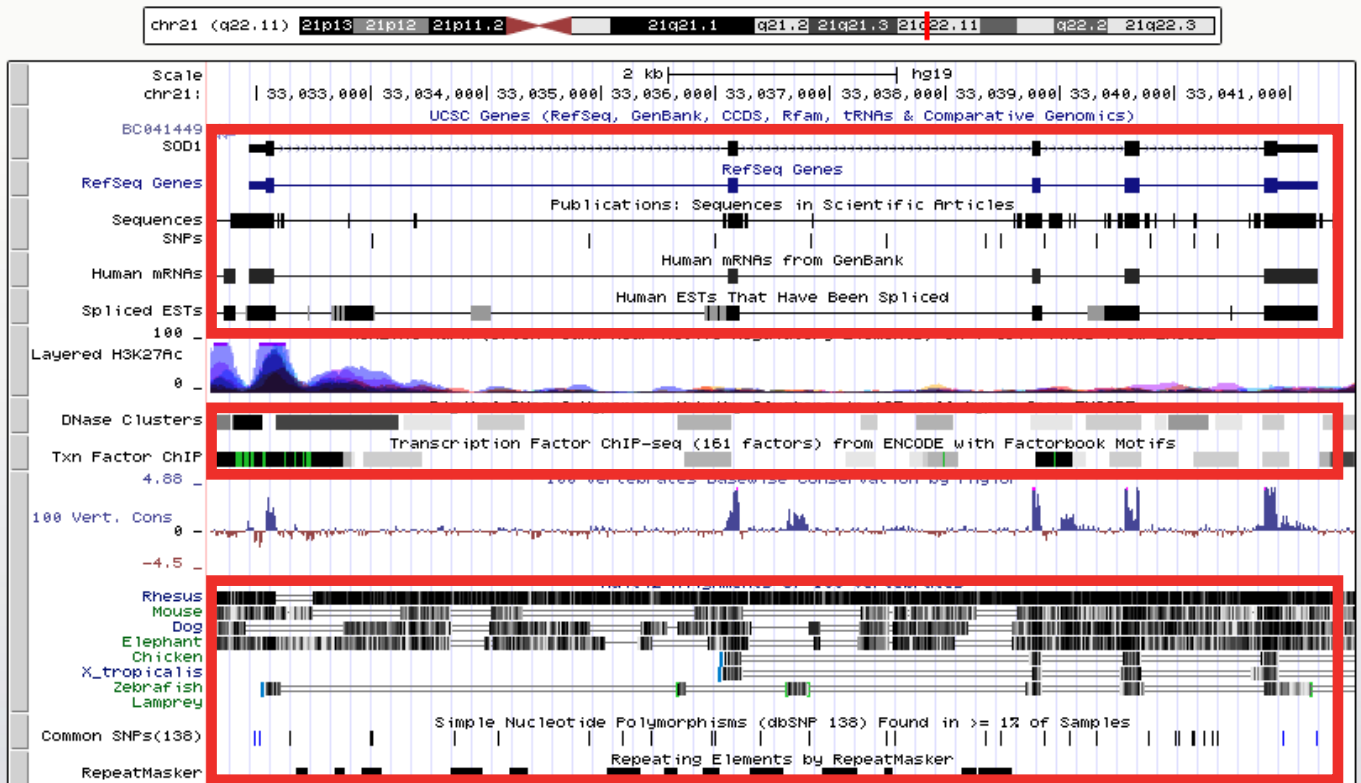
BOOK III.		Page
I.	What Is Lawful in War . . . . .	290
II.	In What Manner the Law of Nations Renders the Property of Subjects Answerable for the Debts of Sovereigns. The Nature of Reptials . . . . .	307
III.	On Just or Solemn War According to the Law of Nations on Declarations of War . . . . .	314

# Annotations:

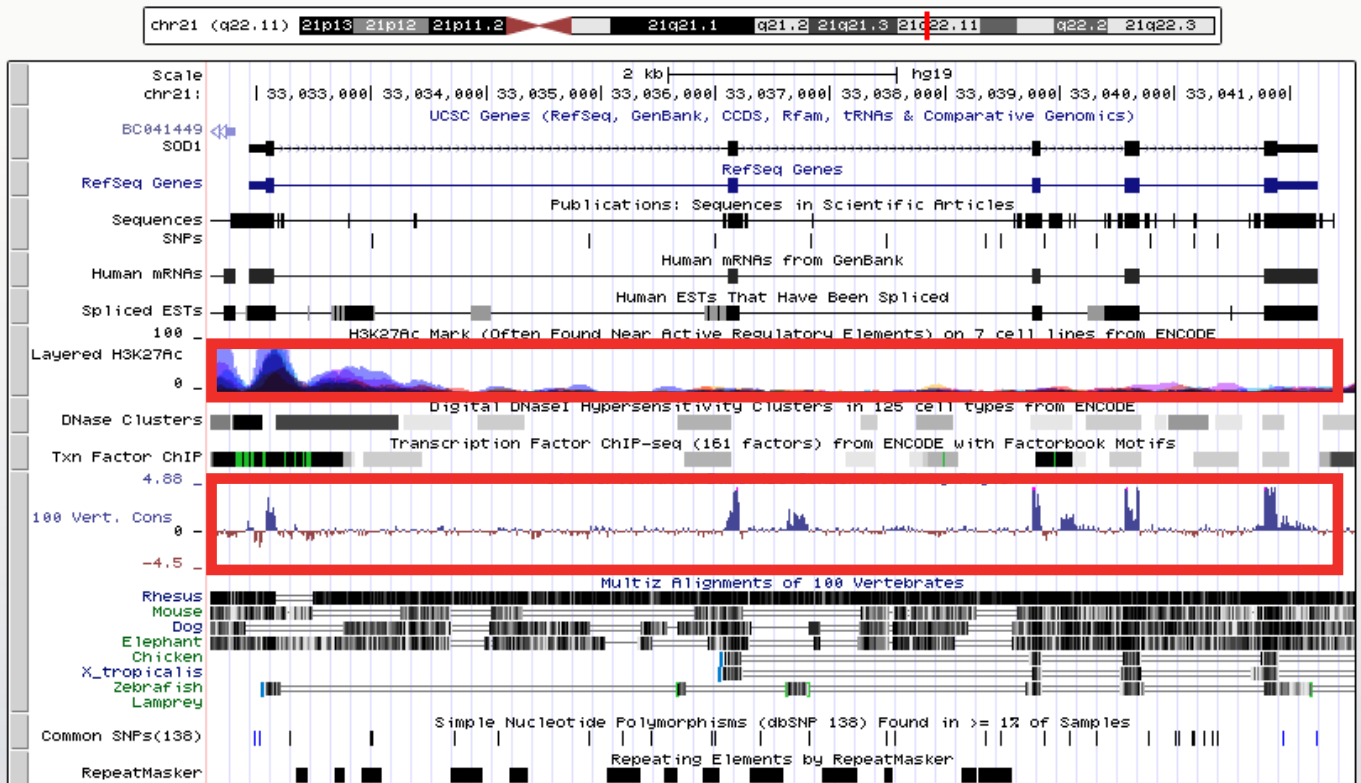




# Annotations: “features” (bed, gff)



# 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

GGAGAAGCaGgAACACGACATCTCCCTCTTCgTCAaGGCTGGCaGtGaTGGGgAAaGCATTGGGAAC  
TGCCcgTTCTCTCAGCGCCTCTTTATGATCCTGTGGCTCAAAGGTGTCATATTTAATGTCACAACCg  
TGGATC

>chr2

GGAGCTCACGCTGATATTTATTGAAAGGATTTGtctCCTGGGAGACTGGGGTGAActtGTTTGGTGG  
TAGCAAGAGAAGCGATAGAATGATTTTTTTGGTTGGGCTTTTTTTttCCTTTCTACCTTCCCCTGCT  
TGTTGAATAGCTGGTGTAGAAAAGGCCCTTTGAGAATACGTGTTCAAGCTTTTGGTTTGTTCTGGG  
GATACTGAC



# Formats: FASTQ

@G59B7NP01B3OX5

AAGCCTTCACCTACACTTTAAGAGCTGAGTAGAATAAGCAAAGTCTAATGCACCAGAAAGAGG...

+

HGGGIIGG999IIHGCGHHIIIIIIIIIIIIIIIIIIIIIEEEIDD99888=A<::=B?==:466...

@G59B7NP01CPQYF

CCATGAGTGGATAGCAGGTTATGGACACAGAGGTTAGAGGAAGCAAGGAAGGAGAAGTAATGAG...

+

IIIIIIIIIIIIIIICCCIIIIIIIIIIIG;;;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



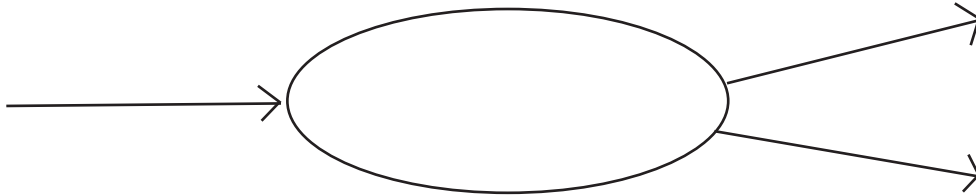
Ken Thompson and Dennis Ritchie

# What is it all about?

- flexibility
- conciseness
- automation

Flexibility: **every** program has

- standard input
- standard output
- standard error



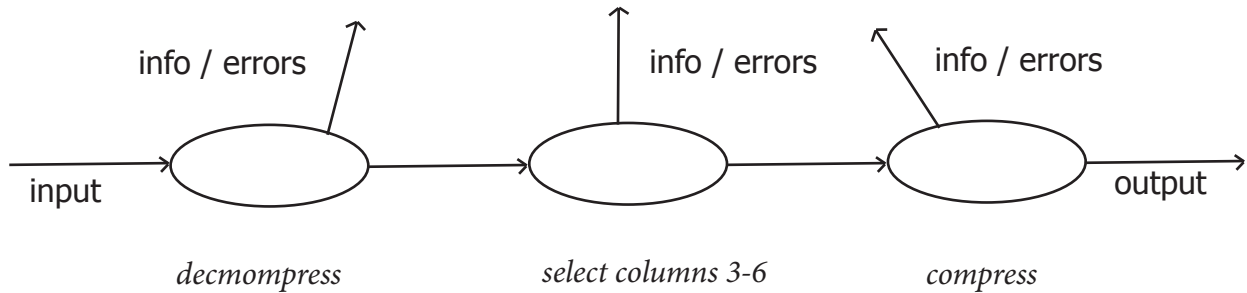
Flexibility: **every** “unix” program

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



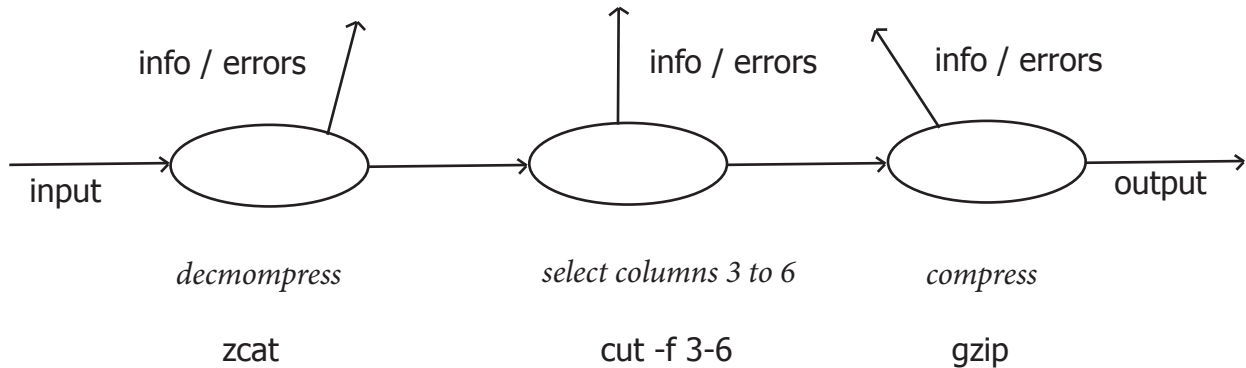
# Flexibility: programs can be chained

USER



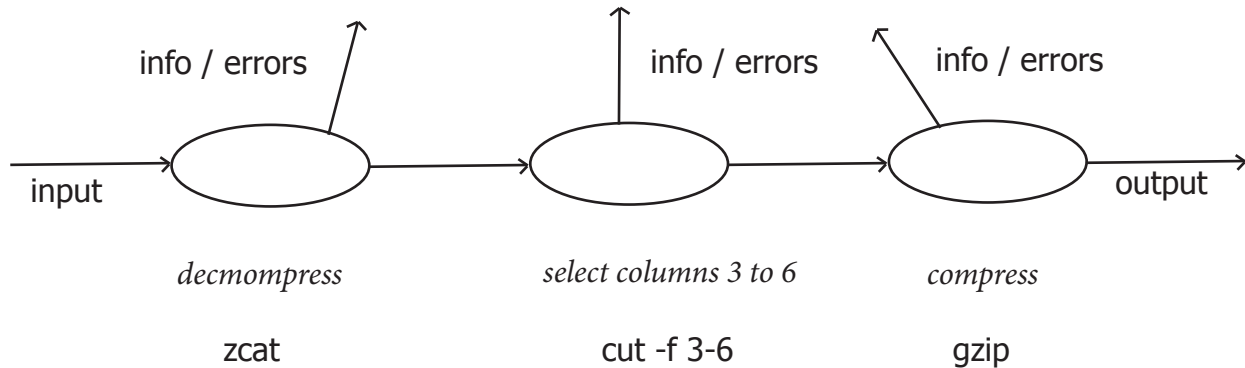
Conciseness: “with just 30 characters you can do almost anything”

USER



Automation: every operation can be stored  
in a text file

USER



```
<bigtable.gz zcat | cut -f 3-6 | gzip >bigtable-3-6.gz
```

And what is Linux then?

