#### SequenceJuxtaposer: Fluid Navigation For Large-Scale Sequence Comparison in Context

#### James Slack<sup>\*</sup>, Kristian Hildebrand<sup>\*†</sup>, Tamara Munzner<sup>\*</sup> and Katherine St.John<sup>¥</sup>

\* University of British Columbia,
 † Bauhaus University Weimar,
 ¥ City University of New York

## Overview

- 1 Introduction
- 1 Previous Work
- 1 Interaction Metaphor
- 1 SequenceJuxtaposer
- 1 Conclusion and Future Work

## Introduction

- Sequence visualization tool
  - 1 Exploration and comparison
- 1 Accordion Drawing
  - 1 Stretch and shrink rubber sheet
  - 1 Borders tacked down



## **Introduction Video**

Zur Anzeige wird der QuickTime™ Dekompressor "MPEG-4 Video" benötigt.



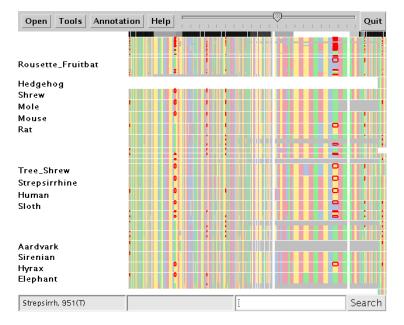


## **Previous Work**

- 1 Accordion Drawing
  - 1 TreeJuxtaposer [Munzner 03]
- 1 Sequence Browsers
  - Ensembl [Hubbard 02], UCSC Genome Browser [Kent 02], NCBI [Wheeler 02]
  - 1 MacClade [Maddison 92], VISTA [Mayor 02], phylo-VISTA [Shah 03]
  - Artemis [Rutherford 00], LalnView [Duret 96], BARD [Spell 03], PipMaker [Schwartz 00]

## **Accordion Drawing**

- SequenceJuxtaposer guarantees 3 key properties
  - 1 Context
  - 1 Visibility
  - 1 Frame Rate



Open Tools Annotatio	n Help		Quit
Rousette_Fruitbat		стстассс	
Hedgehog			
Shrew		GTGTATGCA	
Mole		G T G T A T G C A	
Mouse		G T G T A T G C A	
Rat		G T G T A C G C C	
1			
Too Chann		стстассс	
Tree_Shrew			
Strepsirrhine		<u>стст</u> атссс	
Human		стстатсс с	
Sloth		<u>стстатсс</u>	
-			
Aardvark		N N N N N N N N N	•
Sirenian		GTGTATGCA	
Hyrax		стст 🗛 т с с 🔂	
Elephant		G T A T A T G C A	
Pika, 958(T)		¥	Search

## **Guaranteed Context**

- 1 Focus+Context
  - 1 Combine overview and detail into single view
  - 1 More information visible simultaneously
  - Avoid getting lost while exploring
  - 1 Major information visualization research theme
- 1 Navigation metaphor
  - 1 Rubber sheet with borders tacked down



# **Guaranteed Visibility**

- 1 Highlight marks always visible
  - Never fall outside of current view window
  - 1 Never hidden by something in front
  - Never vanish, even if smaller than one pixel
- 1 Requires efficient algorithms
  - Explicitly checking all items too slow
    Linear in number of pixels, not number of items
  - Details in TreeJuxtaposer paper [Munzner et al, Siggraph03]



# **Guaranteed Frame Rate**

- 1 Need realtime update
  - 1 Focus+Context interaction must be fluid
  - 1 20-30 frames per second
- 1 Computer graphics challenge
  - 1 Progressive rendering



## SequenceJuxtaposer



- 1 Fluid comparison of multiple sequences
- 1 Handles DNA and RNA sequence data
- 1 Provides searching, difference calculation

Open Tools	17	anni	otatio	n 1	Help	in I															_	Qui		
		4.0.1	- dire			19	1	1	1 1	1	1	1	1	1 1	1	1	1	1	1	1 1	1	1	1 1	Qu
Sloth						т								٦	- 1	-	A	С	A	С	A	т		
Armadillo	т	G	ТΤ	A	A	т	т	С	A	т				1	- 7	-	A	с	А	с	А	т		
Anteater	т	A	тτ	G	A	т	т	A	с	т	т			г٦	- 1	-	A	с	A	с	А	т		П
Hedgehog	т	A	тт	т	т	с	т	A	т	т	т			٦	- 1	-	A	с	A	с	A	т		
Mole	т	A	тт	A	G	с	т	G	т		т			٦	- 1	-	A	с	A	с	А	т		
Shrew	т	A	тт	A	G	т	т	G	т	т				٦	- 1	-	A	с	A	С	A	т		Ш
Tenrecid	т	A	тт	A	G	т	т	С	т	т	т			٦	- 1	-	A	т	A	с	A	т		Г
	1							=																
																								Ľ
	f.																							
Human	т	A	ΤТ	A	G	С	т	С	т	т	Т			٦	- 1	-	A	С	А	С	А	т		
					_								-											
Whale	Т	A	тт	A	G	т	т	С	т	т				٦	- 7	-	A	с	А	с	А	т		
Dolphin	т	A	тт	A	G	т	т	С	т	т				٦	- 1	-	A	С	А	с	А	т		
Hippo	т	G	ΤТ	A	А	т	т	F	Т	т	Т			٦	- 1	-	A	С	А	С	А	т		
Llama	Т	A	ΤТ	A	G	т	т	F	с	т	Т			ΤТ	- 1	-	A	С	А	с	А	т		
Ruminant	т	G	тт	A	А	с	т	С	т	т	Т			٦	- 1	-	A	с	А	с	А	т		
Pig	т	A	тт	A	A	т	т	c	т	т	т			٦	- 1	-	A	С	А	с	А	т		
-								_																
	٩.,							=																
Human, 1(T)																							Se	arc

## SequenceJuxtaposer Video

Zur Anzeige wird der QuickTime™ Dekompressor "MPEG-4 Video" benötigt.

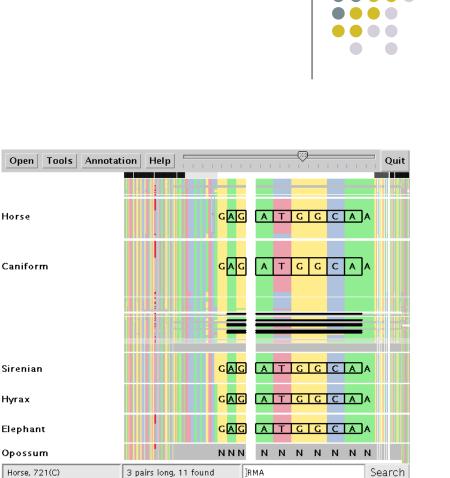


# **Algorithm Complexity**

- 1 Sublinear:
  - 1 Runtime algorithms
- 1 Linear:
  - 1 User-initiated actions
- 1 Subquadratic:
  - 1 Preprocessing algorithms

## Searching

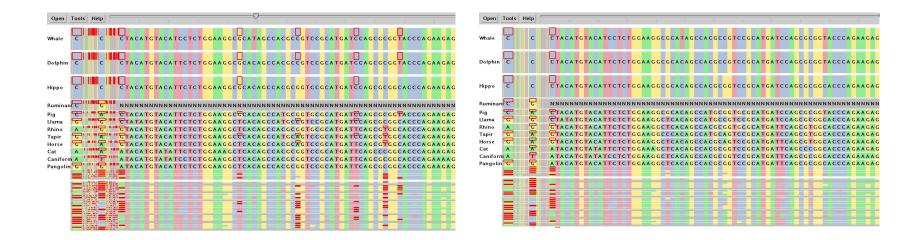
- 1 Search for motifs
  - 1 Protein/Codon search
  - Regular expressions supported
- Results marked with guaranteed visibility



#### Differences



- 1 Explore differences between aligned pairs
  - 1 Slider controls difference threshold in realtime
- 1 Results marked with guaranteed visibility



## Interaction

#### 1 Resizing

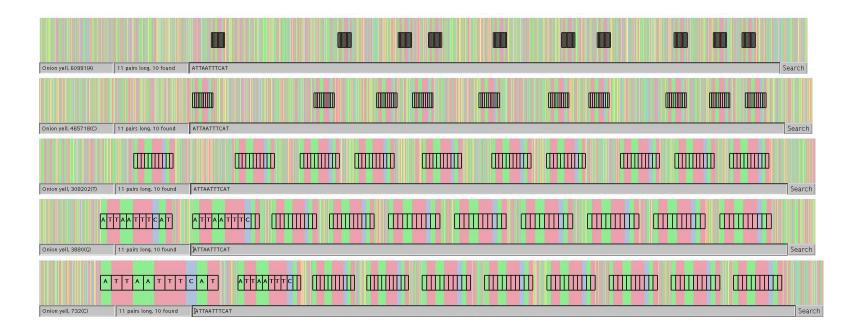
- 1 Expand or contract rectangular areas
- 1 Drag visible rubberband interactively





## Interaction

- 1 Animated transitions
  - 1 Grow and shrink groups
  - 1 Allow user to track visual landmarks





# Drawing



- 1 Very high information density
- 1 Avoid overdrawing in compressed areas
- 1 Progressive rendering
  - 1 Draw for fixed time, check for user interaction
  - Priority queue to draw items in order of current onscreen size



## **Drawing Video**

Zur Anzeige wird der QuickTime™ Dekompressor "MPEG-4 Video" benötigt.

## **Results and Performance**

- 1 Java prototype using OpenGL, GL4Java
- 1 Memory for AD data structures
  - 1 significant, but linear
  - paper: 1.7 million base pairs
  - 1 current: 20 Mbp



## Conclusion



- **1** Accordion Drawing for sequences
  - Powerful new information visualization technique
- 1 Guarantees
  - 1 Context for maintaing orientation
  - 1 Visibility of landmarks: searches, differences
  - 1 Frame rate for realtime response to interaction
- 1 Fluid exploration of big datasets



## **Future Work**

- 1 Performance
  - 1 Memory, speed
- 1 Annotation
- 1 Editing
- 1 Connecting trees and sequences
- 1 Other data types
  - 1 BACs (bacterial artificial chromosomes)

## **Open Source**

- Freely available from <u>http://olduvai.sourceforge.net</u>
  - SequenceJuxtaposer olduvai.sf.net/sj
  - TreeJuxtaposer
    olduvai.sf.net/tj



# Acknowledgements

- 1 Collaboration
  - 1 David Hillis and lab members, UT-Austin
- 1 Discussions
  - 1 Wayne Maddison, David Haussler, Nina Amenta
- 1 Technical writing
  - 1 Ciáran Llachlan Leavitt
- 1 Funding
  - 1 NSF/DEB-0121651/0121682
  - 1 German Academic Exchange Service

