



Ergebnisse des Praktikums zum Modul

“Sequenzanalyse und Genomik”

WS 2010/2011

Universität Leipzig

# Einleitung

- **Deuterostomia:**

- Neumünder
- Urmund wird zum After
- Mund bricht neu durch

- **MiRNA:**

- Kleine RNAs in allen Eukaryoten
- Dienen zur Feinregulation von Genen auf post-transkriptioneller Ebene
- Entstehen durch Prozessierung einer prä-miRNA (Hairpin-Strukturen) mit Drosha/Dicer

- **Let-7:**

- Entdeckt in *Caenorhabditis elegans*
- Steht für lethal (tödlich)
- Evolutionär hochkonserviert
- Steuert zeitliche Entwicklung vieler Tiere

# Aufgabenstellung

- Weitere nicht bekannten let-7-miRNAs in Deuterostomia-Genomen finden
- Erstellen eines gemeinsamen Alignments
- Syntenie der neu gefundenen miRNAs aufklären
- Einordnung zu bekannten let-7-Gruppen
- Erstellen eines Stammbaumes, zum Klären von evolutionären Entwicklungsprozessen und Verwandtschaftsbeziehungen

# Workflow

- Alignment von allen bekannten miRNAs von miRBase der let-7-Familie wurde bereitgestellt
- Vorhandene Precursor-Sequenzen verlängert (Blasten)

```
# let-7a-1
hsa-L-7a-1-1 AUUCACC . CUGGAUGUUCUCUUCACUGUGGGGAGGAGGUAGUAGGUUUGUAUAGUUUUAGGGUCACACCCAC . . . . . CACUGGGAGAUAAUUAACAUCUACU . GUCUUUCCUAAACGUGAUAGAAAAGUCUGCAUCCAGGCGGUCUGAUAG
ptr-L-7a-1-1 AUUCACC . CUGGAUGUUCUCUUCACUGUGGGGAGGAGGUAGUAGGUUUGUAUAGUUUUAGGGUCACACCCAC . . . . . CACUGGGAGAUAAUUAACAUCUACU . GUCUUUCCUAAACGUGAUAGAAAAGUCUGCAUCCAGGCGGUCUGAUAG
ppy-L-7a-1-1 AUUCACC . CUGGAUGUUCUCUUCACUGUGGGGAGGAGGUAGUAGGUUUGUAUAGUUUUAGGGUCACACCCAC . . . . . CACUGGGAGAUAAUUAACAUCUACU . GUCUUUCCUAAACGUGAUAGAAAAGUCUGCAUCCAGGCGGUCUGAUAG
mml-L-7a-1-1 AUUCACC . CUGGAUGUUCUCUUCACUGUGGGGAGGAGGUAGUAGGUUUGUAUAGUUUUAGGGUCACACCCAC . . . . . CACUGGGAGAUAAUUAACAUCUACU . GUCUUUCCUAAACGUGAUAGAAAAGUCUGCAUCCAGGCGGUCUGAUAG
cja-L-7a-1-1_n ATTCACC . CTGGATGTTCTCTTCACTGTGGGATGAGGTAGTAGTTGTATAGTTTTAGGGTCACACCCAC . . . . . CACTGGGAGATAAATAACAATCTACT . GTCTTTCTAACGTGATAGAAAAGTCTGCATCCAGGTGGTCTCATAG
tsy-L-7a-1-1_n GTTCACC . CCGGATGTTCTCTTCACTGTGGGATGAGGTAGTAGTTGTATAGTTTTAGGGTCACACCCAC . . . . . CACTGGGAGATAAATAACAATCTACT . GTCTTTCTAACGTGATAGAAAAGTCTGCATCCAGGCGGCTCATGG
oga-L-7a-1-1_n ATTCAGC . CTGGATGACTCTTCACTGTGGGATGAGGTAGTAGTTGTATAGTTTTAGGGTCACACCCAC . . . . . CACTGGGAGATAAATAACAATCTACT . GTCTTTCTAACGTGATAGAAAAGTCTGCATCCAGGAGGCCTCATGC
mim-L-7a-1-1_n ATTCACC . CTGGATGACTCTTCACTGTGGGATGAGGTAGTAGTTGTATAGTTTTAGGGTCACACCCAC . . . . . CACTGGGAGATAAATAACAATCTACT . ATCTACT . GTCTTTCTAACGTGATAGAAAAGTCTGCATCCAGGTGGCTCATAG
tbe-L-7a-1-1_n CAATATTCACCCCTAGTTCTTTTCACTGTAGGATAAGGTAGTAGATTGTATAGTTCTAGGGTCACACCCAT . . . . . CATTGGAAGATCACTCTATAACAATCTACTGTCTTTCTCAGATGGTAGAAAAGATCGGCNNNNNNNNNNNNNNNNNN
mmu-L-7a-1-1 AUUCACC . CUGGAUGUUCUCUUCACUGUGGGGAGGAGGUAGUAGGUUUGUAUAGUUUUAGGGUCACACCCAC . . . . . CACUGGGAGAUAAUUAACAUCUACU . GUCUUUCCUAAAGGUGAUGGAAAAGUCUGCAUUCUAGGGGUCUAG
rno-L-7a-1-1 AUUUGCC . CUGGAUGUUGUCUUCACUGUGGGGAGGAGGUAGUAGGUUUGUAUAGUUUUAGGGUCACACCCAC . . . . . CACUGGGAGAUAAUUAACAUCUACU . GUCUUUCCUAAAGGUGAUGGAAAAGUCUGCAUUCGUG . GAUCUAGUAG
bta-L-7a-1-1 AUUCUCC . CUGGACAUUCUCUUCACUGUGGGGAGGAGGUAGUAGGUUUGUAUAGUUUUAGGGUCACACCCAC . . . . . CACUGGGAGAUAAUUAACAUCUACU . GUCUUUCCUAAAGGUGAUGAAGAAAAGUCUGCAUCCAAACAACUCACAC
mdo-L-7a-1-1 AUCUACAACUAAAACAUAUCUUCACUGUGGGGAGGAGGUAGUAGGUUUGUAUAGUUUUAGGGUCACACCCAC . . . . . CACUGGGAGAUAAUUAACAUCUACU . GUCUUUCCUAAAGUG . GGAGAAAUCUCCAGGAGUUAUCUUU . UUAU
dre-L-7a-1-1 GUGGACUGG . AGGUUACAUUGUACGGUGGGGAGGAGGUAGUAGGUUUGUAUAGUUUUAGGGUCACACCCAC . . . . . ACUGGGAGAUAAUUAACAUCUACU . GUCUUUCUCAAAGUC . CAAUUCUACAGUACAGAGUUUCAUCCU . GUAC
gga-L-7a-1-1 ACAUACCCUCUGAAGAUCCUGCACUGUGGGGAGGAGGUAGUAGGUUUGUAUAGUUUUAGGGUCAUACCCGC . . . . . AACUGGGAGAUAAUUAACAUCUACU . GUCUUUCCUAAAGCAGCAGAAAUAACAACGGAGACCGUUCAGUUU
tgu-L-7a-1-1 GCCCCAGCUCUGAAGAUCCUGCACUGUGGGGAGGAGGUAGUAGGUUUGUAUAGUUUUAGGGUCAUACCCAC . . . . . CACUGGGAGAUAAUUAACAUCUACU . GUCUUUCCUAAAGCAGCAGAAAUAACAACUAGGAGCUCUCAGUUU
xtr-L-7a-1-1 AGCUGCUCUCUGGCCCCACCCGCCUG . UGGGAUGAGGUAGUAGGUUUGUAUAGUUUUAGGGCCGCACCCAC . . . . . CAUUGGGAGAGAUAACAACAUCUACU . GUCUCUCCUAAAGUGGCCGGAUUCUCCUACGACCAUUGUCUUUAG
```

# Workflow

- blasten aller bekannter mir-let-7 RNAs gegen die Genome von Mitgliedern der Deuterostomier
- ordnen des Blastoutputs mittels R-Skript nach Locus und E-Value

1	147	xtr-L-7x-c-1	scaffold_159	100	147	0	0	1	147	329172	329026	2.6e-77	291.9
1	147	tgu-L-7g-1	scaffold_159	89.36	94	10	0	29	122	329144	329051	8.1e-22	107.5
1	147	tgu-L-7g-2	scaffold_159	89.36	94	10	0	29	122	329144	329051	8.1e-22	107.5
1	147	gga-L-7g-1	scaffold_159	89.25	93	10	0	29	121	329144	329052	3.2e-21	105.6
1	147	mim-L-7g-1_n	scaffold_159	90.24	82	8	0	29	110	329144	329063	2e-19	99.61
1	147	cpo-L-7g-1_n	scaffold_159	90.24	82	8	0	29	110	329144	329063	2e-19	99.61
1	147	mdo-L-7g-1	scaffold_159	88.17	93	11	0	29	121	329144	329052	7.8e-19	97.63
1	147	fca-L-7g-1_n	scaffold_159	89.02	82	9	0	28	109	329144	329063	4.8e-17	91.68
1	147	oan-L-7g-1	scaffold_159	87.23	94	12	0	29	122	329144	329051	4.8e-17	91.68
1	147	eeu-L-7g-1_n	scaffold_159	89.02	82	9	0	29	110	329144	329063	4.9e-17	91.68
1	147	pva-L-7g-1_n	scaffold_159	89.02	82	9	0	29	110	329144	329063	4.9e-17	91.68
1	147	sar-L-7g-1_n	scaffold_159	89.02	82	9	0	29	110	329144	329063	4.9e-17	91.68
1	147	tbe-L-7g-1_n	scaffold_159	89.02	82	9	0	29	110	329144	329063	4.9e-17	91.68
1	147	bta-L-7g-1	scaffold_159	89.02	82	9	0	29	110	329144	329063	4.9e-17	91.68
1	147	laf-L-7g-1_n	scaffold_159	89.02	82	9	0	29	110	329144	329063	4.9e-17	91.68
1	147	ttr-L-7g-2_n	scaffold_159	89.02	82	9	0	29	110	329144	329063	4.9e-17	91.68
1	147	ssc-L-7g-1	scaffold_159	89.02	82	9	0	29	110	329144	329063	4.9e-17	91.68
1	147	ttr-L-7g-1_n	scaffold_159	89.02	82	9	0	29	110	329144	329063	4.9e-17	91.68
1	147	ocu-L-7g-1_n	scaffold_159	89.02	82	9	0	29	110	329144	329063	4.9e-17	91.68
1	147	mlu-L-7g-1_n	scaffold_159	89.02	82	9	0	29	110	329144	329063	4.9e-17	91.68
1	147	tsy-L-7g-1_n	scaffold_159	89.02	82	9	0	29	110	329144	329063	4.9e-17	91.68
1	147	ptr-L-7g-1	scaffold_159	89.02	82	9	0	29	110	329144	329063	4.9e-17	91.68
1	147	mml-L-7g-1	scaffold_159	89.02	82	9	0	29	110	329144	329063	4.9e-17	91.68
1	147	cja-L-7g-1_n	scaffold_159	89.02	82	9	0	29	110	329144	329063	4.9e-17	91.68
1	147	aml-L-7g-1_n	scaffold_159	89.02	82	9	0	29	110	329144	329063	4.9e-17	91.68
1	147	dno-L-7g-1_n	scaffold_159	89.02	82	9	0	29	110	329144	329063	4.9e-17	91.68
1	147	vpa-L-7g-1_n	scaffold_159	89.02	82	9	0	29	110	329144	329063	4.9e-17	91.68
1	147	eca-L-7g-1	scaffold_159	89.02	82	9	0	29	110	329144	329063	4.9e-17	91.68
1	147	mmu-L-7g-1	scaffold_159	89.02	82	9	0	29	110	329144	329063	4.9e-17	91.68
1	147	cfa-L-7g-1	scaffold_159	89.02	82	9	0	29	110	329144	329063	4.9e-17	91.68
1	147	ppy-L-7g-1	scaffold_159	89.02	82	9	0	29	110	329144	329063	4.9e-17	91.68
1	147	hsa-L-7g-1	scaffold_159	89.02	82	9	0	29	110	329144	329063	4.9e-17	91.68
1	147	opr-L-7g-1_n	scaffold_159	89.02	82	9	0	29	110	329144	329063	4.9e-17	91.68
1	147	oar-L-7g-1_n	scaffold_159	89.02	82	8	1	29	109	329144	329063	1.2e-14	83.75
1	147	dor-L-7g-1_n	scaffold_159	87.8	82	10	0	29	110	329144	329063	1.2e-14	83.75
1	147	oga-L-7g-1_n	scaffold_159	88.24	68	8	0	29	96	329144	329077	4.4e-11	71.86
1	147	ola-L-7x-c-n	scaffold_159	97.22	36	1	0	31	66	329142	329107	1.1e-08	63.93
1	147	gac-L-7x-c-n	scaffold_159	95	40	2	0	31	64	329148	329109	1.1e-08	63.93
1	147	tni-L-7x-c-1	scaffold_159	94.44	36	2	0	31	66	329142	329107	2.6e-06	56
1	147	fru-L-7x-c-1	scaffold_159	94.44	36	2	0	30	65	329142	329107	2.7e-06	56
1	147	drg-L-7x-c-1	scaffold_159	96.43	38	1	0	27	54	329146	329119	0.00066	48.07
1	147	ete-L-7g-1_n	scaffold_159	83.82	68	11	0	29	96	329144	329077	0.00067	48.07
2	146	xtr-L-mir-98-1	scaffold_507	100	146	0	0	1	146	538561	538416	1e-76	289.9
2	146	oan-L-mir-98-1	scaffold_507	100	31	0	0	29	59	538534	538504	4.3e-08	61.95
2	146	str-L-mir-98-1_n	scaffold_507	97.06	34	1	0	27	60	538537	538504	1.7e-07	59.96
2	146	eeu-L-mir-98-1n	scaffold_507	97.06	34	1	0	24	57	538537	538504	1.7e-07	59.96
2	146	sar-L-mir-98-1n	scaffold_507	96.97	33	1	0	23	55	538536	538504	6.5e-07	57.98
2	146	eca-L-mir-98-1	scaffold_507	96.97	33	1	0	25	57	538536	538504	6.7e-07	57.98
2	146	ppy-L-mir-98-1	scaffold_507	96.97	33	1	0	25	57	538536	538504	6.7e-07	57.98
2	146	mml-L-mir-98-1	scaffold_507	96.97	33	1	0	25	57	538536	538504	6.7e-07	57.98
2	146	mim-L-mir-98-1_n	scaffold_507	96.97	33	1	0	25	57	538536	538504	6.7e-07	57.98

# Workflow

- Erstellen von Coordinates-Files aller Arten
- Ableiten der Loci an Vorbild des Menschen und Huhns

```

Procavia capensis (pca) miRNAs
=====

```

MI-id	mir-name	chr	position	strand	mat(annotated)	Remarks
	pca-L-7c-1_n	scaffold_12582	10456 10592	+		
	pca-L-7i-1_n	scaffold_101430	369 499	+		
	pca-L-7e-1_n	scaffold_21308	31849 31966	-		█
	pca-L-7a-1-1_n	scaffold_253307	962 1094	-		
	pca-L-7a-3-1_n	scaffold_274020	444 560	-		
	pca-L-7f-1-1_n	scaffold_253307	583 710	-		
	pca-L-7f-2-1_n	GeneScaffold_934	45220 45358	-		
	pca-L-7a-2-1_n	scaffold_50800	4325 4453	+		
	pca-L-mir-98-1	GeneScaffold_934	44180 44304	-		
	pca-L-7x-f-1	scaffold_253307	583 693	+		
	mir-100	scaffold_53364	12647 12722	+		
	mir-99b	scaffold_21308	32053 32122	-		
	mir-125a	scaffold_21308	31294 31378	-		
	mir-99a	scaffold_12582	9733 9811	+		

#Loci	chr	mir1	abstand	mir2	abstand	mir3
I						
II	scaffold_253307	pca-L-7a-1-1_n	[84]	pca-L-7f-1-1_n		
III	scaffold_50800	pca-L-7a-2-1_n				
IV	scaffold_101430	pca-L-7i-1_n				
V	scaffold_21308	pca-mir-99b	[87]	pca-L-7e-1_n	[471]	pca-mir-125a
VI	scaffold_12582	pca-mir-99a	[645]	pca-L-7c-1_n		
VII	scaffold_274020	pca-L-7a-3-1_n				
VIII	GeneScaffold_934	pca-L-7f-2-1_n	[916]	pca-L-mir-98-1		
IX	scaffold_253307	pca-L-7x-f-1				



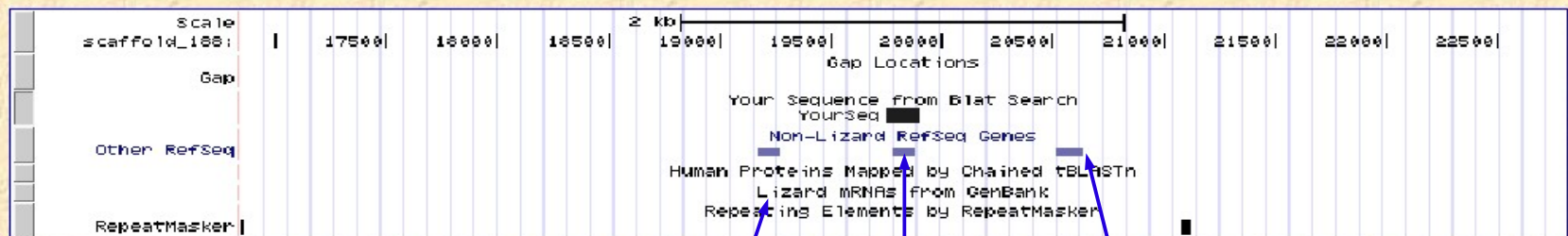
# Workflow

- Erstellen eines neuen Alignments aus den bereits annotierten und den neu gefundenen miRNAs mittels ClustalW
- Unterteilung des multiplen Alignments in Untergruppen der let-7-miRNA-Familie
- Überprüfen und korrigieren der erstellten Gruppen durch Annotation von Syntenieinformationen

## UCSC Genome Browser on Lizard Feb. 2007 (Broad/anoCar1) Assembly

move <<< << < > >> >>> zoom in 1.5x 3x 10x base zoom out 1.5x 3x 10x

position/search scaffold\_188:16,848-22,801 jump clear size 5,954 bp. configure



let-7a-1 acr-L-7f-1-1\_n let-7d



# Workflow

- Erstellen von Syntenie-Files für alle eingeteilten Blocks

```
# let-7f-1
10 let-7a-1 - 280 hsa-L-7f-1-1 2300 let-7d
10 let-7a-1 - 620 oan-L-7f-1-1 1200 let-7d
10 let-7a-1 - 630 mdo-L-7f-1-1 4700 let-7d
10 let-7a-1 - 200 laf-L-7f-1-1_n 13000 let-7d
10 let-7a-1 - 287 eca-L-7f-1-1 1800 let-7d
10 let-7a-1 - 260 cfa-L-7f-1-1 1700 let-7d
10 let-7a-1 - 205 rno-L-7f-1-1 1600 let-7d
10 let-7a-1 - 500 acr-L-7f-1-1_n 600 let-7d
10 let-7a-1 - 309 gga-L-7f-1-1 860 let-7d
10 let-7a-1 - 70 dre-L-7f-1-1 2600 Wnt7a
```

# Workflow

- Erstellen von Syntenie-Files für alle eingeteilten Blocks für 30 Arten, deren Genome im UCSC-Browser vorhanden waren

```
# let-7f-1
10 let-7a-1 - 280 hsa-L-7f-1-1 2300 let-7d
10 let-7a-1 - 620 oan-L-7f-1-1 1200 let-7d
10 let-7a-1 - 630 mdo-L-7f-1-1 4700 let-7d
10 let-7a-1 - 200 laf-L-7f-1-1_n 13000 let-7d
10 let-7a-1 - 287 eca-L-7f-1-1 1800 let-7d
10 let-7a-1 - 260 cfa-L-7f-1-1 1700 let-7d
10 let-7a-1 - 205 rno-L-7f-1-1 1600 let-7d
10 let-7a-1 - 500 acr-L-7f-1-1_n 600 let-7d
10 let-7a-1 - 309 gga-L-7f-1-1 860 let-7d
10 let-7a-1 - 70 dre-L-7f-1-1 2600 Wnt7a
```

- Erstellen eines gemeinsamen Syntenie-Files
- Automatisiertes Ordnen des Files (Jana) um nicht korrekt einsortierte miRNAs in die richtigen Gruppen einzusortieren

# Ergebnisse und Diskussion

- Viele neue miRNAs in verschiedenen Spezies gefunden
- Durch Syntenieinformation und daraus resultierende Ordnung in Untergruppen Grundlage für einheitliche Namensgebung geschaffen
- insbesondere L-7x-f Gruppe enorm erweitert und in Untergruppen eingeteilt --> bedarf genauerer Analyse, konnte im Rahmen des Praktikums nicht erledigt werden
- To do:
  - Vervollständigen der Syntenieinformation
  - Anschließend erstellen von phylogenetischen Bäumen
  - Daraus ableiten von Verwandtschaftsbeziehungen und der evolutionären Entwicklung der let-7-miRNA-Familie

Wir laden nun herzlich zur Diskussion und weiteren Ergebnisfindung ein!