

Использование КС-грамматики для распознавания 16s рРНК

Семён Григорьев, Дмитрий Ковалёв

13 сентября 2017 г.

1 Введение

Одна из задач, возникающих в биоинформатике — задача поиска и классификации маркерных цепочек, используемых для обнаружения и классификации организмов. При этом, многие маркерные цепочки обладают достаточно богатой вторичной структурой, что позволяет сделать предположение, что её можно использовать для решения данной задачи. Более того, известно, что некоторые участки обладают достаточно консервативной вторичной структурой, например, центральный домен 16s [10]. Использование вторичной структуры при решении задач поиска и классификации цепочек не является новой идеей. Она используется как в широко распространённых инструментах (например, Infernal [1]), так и во многих теоретических исследованиях [2, 3, 4, 5, 6].

Вторичная структура цепочек может быть описана с помощью грамматик. В зависимости от требуемой точности могут использоваться разные классы грамматик: регулярные, контекстно-свободные, конъюнктивные. В данной работе будут использоваться контекстно-свободные грамматики и соответствующий класс языков. В результате, по аналогии с поиском по регулярному выражению, задача поиска сведётся к поиску структурного шаблона, заданного контекстно-свободной грамматикой. Так как вторичная структура больших цепочек может быть достаточно сложной, соответствующая грамматика также оказывается сложной. При этом часто необходимо искать баланс между сложностью грамматики, которая непосредственно связана с детализацией описания вторичной структуры, и производительностью и точностью поиска.

Одна из маркерных цепочек, часто используемая в настоящее время — это 16s rRNA. Данный отчёт описывает эксперимент по распознаванию 16s с использованием информации только о её вторичной структуре, описанной контекстно-свободной грамматикой.

2 Описание вторичной структуры с помощью грамматики

Для спецификации грамматики был использован язык YARD [9], основанный на ECFG [7] с различными расширениями. В правых частях можно использовать конструкции регулярных выражений.

В таблице 1 представлены основные конструкции и примитивы, использовавшиеся при написании грамматики. Так как описание несовпадений в стеке в общем случае является сложной задачей (возможно, неразрешимой в терминах контекстно-свободных грамматик), были использованы такие правила, как $stem_el<s>$, описывающие приближение множества стеков с несовпадениями.

Грамматическая конструкция	Описание
any	Один из нуклеотидов A, U, C, G .
$any^*[n..m]$	Цепочка нуклеотидов длины от n до m .
$stemN<s>$	Стем высоты N со свободной частью s (последовательность любых конструкций грамматики).
$mk_stem<s>$	Стем произвольной высоты (от 0 до N) со свободной частью s .
$stem_el<s>$	Стем, позволяющий одно несовпадение, при этом требующий, чтобы подряд было не менее двух парных элементов.

Таблица 1: Базовые конструкции грамматики

$stem4<any^*[3..5]>$	$mk_stem<any^*[1..2] stem2<any^*[3..4]> any^*[2..5]>$
<pre> A C U G G — C A — U G — C G — C </pre>	<pre> C A G — G C — U A — G C — A A C U — A G — C A — U G — C G — C </pre>

Таблица 2: Примеры описания структур

3 Эксперименты

Для проведения эксперимента прежде всего необходима контекстно-свободная грамматика, описывающая вторичную структуру искомой цепочки. Построение грамматики, задающей вторичную структуру, в настоящий момент выполняется вручную, однако возможен и вывод грамматики, но это тема для отдельного исследования.

Используемая в эксперименте грамматика, которая была построена нами, приведена в приложении А. В качестве образца была использована эталонная вторичная структура 16s *Escherichia coli*. Терминальный алфавит состоит из четырех символов-нуклеотидов: *A, U, C, G*. Для спецификации стемов активно используются параметризуемые правила, что позволяет сделать грамматику достаточно компактной. Ключевое слово `inline` является служебным и используется для того, чтобы подсказать генератору синтаксических анализаторов, как именно преобразовывать грамматику в ходе работы.

Всего было поставлено два эксперимента: обработка базы известных последовательностей 16s и поиск 16s в полноразмерных размеченных геномах. Целью первого эксперимента была оценка качества составленной нами грамматики на предмет полноты детектирования цепочек: вычислялся процент нераспознанных цепочек относительно всех обработанных. Цель второго — оценка количества ложных срабатываний. Вместе с этим оценивалась и полнота поиска.

Для первого эксперимента была использована база цепочек проекта SILVA [8], которая содержит более 20 тысяч различных последовательностей. Результаты данного эксперимента приведены в таблице 3, где '+' — количество распознанных цепочек для соответствующего домена, '-' — количество нераспознанных. В итоге, при использовании грамматики для центрального домена распознано 98.16% цепочек, являющихся 16s рРНК, а при использовании грамматики для 5'М домена — 63.13%. Кроме этого, можно заметить, что 5'М домен лучше разделяет домены.

Домен	Стартовый нетерминал	Бактерии		Эукариоты		Археи	
		+	-	+	-	+	-
Центральный	h19	17878	335	2153	3165	306	13
5'М	h3	11498	6715	64	5254	81	238

Таблица 3: Результаты анализа базы организмов

Для второго эксперимента использовались 100 размеченных геномов из базы NCBI. Поиск по центральному домену показал очень большое количество ложных срабатываний, поэтому результаты здесь приводиться не будут. Часть результатов поиска по 5'М домену приведены в таблице 4. В таблице указан код организма в NCBI, его название, количество размеченных и правильно обнаруженных последовательностей 16s (Expected и Covered соответственно), количество ложных срабатываний (FP-intervals), средняя длина и отклонение для цепочек, соответствующих ложным срабатываниям. Последние два столбца могут помочь оценить объём работы по дополнительной фильтрации ложных срабатываний. В итоге, для 100 геномов получены следующие результаты:

- среднее количество ложных срабатываний на геном равно 299, отклонение равно 221,54;
- средняя доля правильно обнаруженных цепочек равна 0,98 при отклонении 0,11.

Также, в таблице 5 приведены результаты поиска по совмещённой грамматике для 5'М и центрального домена. Обращает на себя резкое снижение количества ложных срабатываний. При этом, однако, уменьшается и количество положительных совпадений.

NCBI ID	Name	Expected	Covered	FP-intervals	Length(avg.)	Length SD
NZ_CP012959.1	Aggregatibacter actinomycetemcomitans strain 624	6	6	164	614.6	246.4
NC_014640.1	Achromobacter xylosoxidans A8	3	3	556	586.9	244.9
NC_005966.1	Acinetobacter sp. ADP1 complete genome	7	7	211	552.8	134.4
NZ_CP009448.1	Achromobacter xylosoxidans C54	3	3	598	579.1	201.6
NC_013171.1	Anaerococcus prevotii DSM 20548	4	3	245	584.6	256.0
NZ_CP012590.1	Actinomyces sp. oral taxon 414 strain F0588	3	3	363	591.5	196.8
NZ_CP014060.1	Achromobacter xylosoxidans strain FDAARGOS_147	3	0	741	614.1	266.0
NZ_CP007502.1	Aggregatibacter actinomycetemcomitans HK1651	6	6	113	569.8	173.6
NZ_LT635457.1	Actinomyces sp. Marseille-P2985	3	3	71	570.0	153.4
NZ_CP012046.1	Achromobacter xylosoxidans strain MN001	3	3	560	599.9	222.9
NZ_LN831029.1	Achromobacter xylosoxidans genome assembly NCTC10807	3	3	653	582.5	206.0
NZ_CP015594.1	Acinetobacter sp. NCu2D-2	7	7	125	544.1	117.1
NC_012913.1	Aggregatibacter aphrophilus NJ8700	6	6	157	551.2	147.7
NZ_CP020468.1	Actinomyces sp. pika_114	3	3	103	664.1	596.9
NZ_CP014232.1	Actinomyces oris strain T14V	3	3	200	606.1	227.7
NZ_CP015110.1	Acinetobacter sp. TGL-Y2	7	7	234	561.9	173.1
NZ_CP012608.1	Acinetobacter sp. TTH0-4	7	7	204	575.3	169.6
NZ_CP017812.1	Actinomyces sp. VUL4_3	3	3	125	662.4	326.1
NZ_CP012067.1	Aggregatibacter aphrophilus strain W10433	6	6	162	551.3	119.0
NZ_CP012072.1	Actinomyces meyeri strain W712	3	3	197	631.5	313.8
NC_000964.3	Bacillus subtilis subsp. subtilis str. 168 chromosome	10	10	278	558.2	154.3
NZ_CP016852.1	Bacillus subtilis subsp. subtilis strain 168G	10	10	278	558.9	154.3
NZ_CP009902.1	Bacillus anthracis strain 2002013094	11	10	1352	656.2	269.3
NZ_CP017763.1	Bacillus subtilis strain 29R7-12	10	10	297	538.0	110.4
NZ_CP010314.1	Bacillus subtilis subsp. subtilis strain 3NA	10	10	273	556.3	150.7
NC_012659.1	Bacillus anthracis str. A0248	11	11	1302	660.8	285.9
NC_011835.1	Bifidobacterium animalis subsp. lactis AD011	2	2	190	604.0	223.0
NZ_CP009748.1	Bacillus subtilis strain ATCC 13952	7	7	216	553.3	147.2
NZ_CP009749.1	Bacillus subtilis strain ATCC 19217	7	7	231	542.8	147.4
NC_017834.1	Bifidobacterium animalis subsp. animalis ATCC 25527	4	4	157	653.2	299.0
NC_022523.1	Bifidobacterium animalis subsp. lactis ATCC 27673	4	4	176	613.7	252.4
NC_017866.1	Bifidobacterium animalis subsp. lactis B420	4	4	192	605.1	243.0
NZ_CP014227.1	Capnocytophaga haemolytica strain CCUG 32990	4	4	405	681.3	312.6
NZ_CP014230.1	Desulfomicrobium orale DSM 12838	2	2	198	566.4	162.8
NC_013162.1	Capnocytophaga ochracea DSM 7271	4	4	161	613.1	184.1
NZ_CP012475.1	Bacillus clausii strain ENTPro	7	7	350	555.3	147.7
NZ_CP017037.1	Dialister pneumosintes strain F0677	5	5	407	647.0	257.2
NZ_CP012366.1	Enterococcus durans strain KLDS6.0933	6	6	202	532.3	101.7

NZ_CP012384.1	Enterococcus durans strain KLDS 6.0930	6	6	203	532.0	101.4
NC_006582.1	Bacillus clausii KSM-K16 DNA	7	7	369	559.0	176.2
NZ_CP016923.1	Klebsiella pneumoniae isolate 11	8	8	381	543.7	150.8
NZ_CP008740.1	Haemophilus influenzae 2019	6	6	132	545.0	107.4
NZ_CP016926.1	Klebsiella pneumoniae isolate 23	8	8	389	553.2	162.1
NZ_CP011313.1	Klebsiella pneumoniae subsp. pneumoniae strain 234-12	8	8	392	556.3	154.4
NC_013721.1	Gardnerella vaginalis 409-05	2	2	237	601.6	229.9
NZ_CP007470.1	Haemophilus influenzae strain 477	6	6	136	560.1	118.2
NZ_CP007472.1	Haemophilus influenzae strain 723	6	6	146	564.1	152.6
NC_007146.2	Haemophilus influenzae 86-028NP	6	6	157	548.1	104.8
NC_014644.1	Gardnerella vaginalis ATCC 14019 chromosome	2	2	267	597.4	190.2
NC_003454.1	Fusobacterium nucleatum subsp. nucleatum ATCC 25586	5	5	766	654.9	281.7
NC_010376.1	Fingoldia magna ATCC 29328 DNA	4	4	339	645.8	335.1
NC_018610.1	Lactobacillus buchneri CD034	5	5	107	533.4	173.3
NC_000908.2	Mycoplasma genitalium G37	1	1	30	533.8	98.9
NZ_CP019058.1	Gardnerella vaginalis strain GV37	2	2	282	588.6	184.2
NZ_CP012716.1	Fusobacterium nucleatum subsp. nucleatum ChDC F3162	4	4	819	655.7	301.7
NC_018498.1	Mycoplasma genitalium M2288	1	1	32	519.3	92.5
NC_018496.1	Mycoplasma genitalium M6282	1	1	30	530.3	95.6
NC_015428.1	Lactobacillus buchneri NRRL B-30929	5	5	97	524.6	122.8
NZ_CP019323.1	Lactobacillus sp. WiKim39	4	4	443	605.1	206.9
NZ_CP009531.1	Lactobacillus sp. wkB8	4	4	108	555.3	162.9
NC_014752.1	Neisseria lactamica 020-06 complete genome	4	4	176	643.8	294.6
NZ_CP013696.1	Pseudomonas aeruginosa strain 12-4-4(59)	4	4	374	555.6	160.0
NC_017534.1	Propionibacterium acnes 266	3	3	197	605.1	270.4
NZ_CP012889.1	Porphyromonas gingivalis 381	4	4	105	520.1	78.0
NC_017535.1	Propionibacterium acnes 6609	3	3	195	609.4	279.5
NZ_AP014839.1	Pseudomonas aeruginosa DNA	4	4	377	570.3	181.0
NZ_CP011995.1	Porphyromonas gingivalis strain A7436	4	4	98	540.6	151.9
NZ_CP013680.1	Pseudomonas aeruginosa AES-1R	1	1	369	556.4	137.9
NZ_CP015347.1	Proteus mirabilis strain AOUC-001	7	7	261	550.1	176.0
NZ_CP020052.1	Proteus mirabilis strain AR_0059	7	7	259	557.6	152.2
NZ_CP021694.1	Proteus mirabilis strain AR_0155	7	7	281	565.3	171.6
NC_017550.1	Propionibacterium acnes ATCC 11828	2	2	183	583.6	210.5
NZ_CP017149.1	Pseudomonas aeruginosa strain ATCC 15692	4	4	377	556.7	159.6
NZ_CP007726.1	Neisseria elongata subsp. glycolytica ATCC 29315	4	4	179	655.0	324.3
NC_010729.1	Porphyromonas gingivalis ATCC 33277 DNA	4	4	103	525.3	104.1
NC_018707.1	Propionibacterium acnes C1	3	3	206	598.5	265.6
NZ_CP012830.1	Pseudomonas fluorescens strain FW300-N2E3	6	6	433	564.4	158.7
NZ_CP011117.1	Pseudomonas fluorescens strain LBUM223	6	6	437	591.4	217.4
NZ_CP012073.1	Ottowia sp. oral taxon 894 strain W10237	3	3	269	636.5	258.7

NZ_CP019894.1	Neisseria lactamica strain Y92-1009	4	4	177	603.6	195.8
NZ_CP007241.1	Streptococcus pyogenes strain 1E1	6	6	126	582.6	180.2
NZ_CP016756.1	Stenotrophomonas nitritireducens strain 2001	4	4	595	616.9	252.8
NC_014498.1	Streptococcus pneumoniae 670-6B	4	4	257	591.7	199.0
NC_012468.1	Streptococcus pneumoniae 70585	4	4	246	598.4	201.2
NC_002967.9	Treponema denticola ATCC 35405 chromosome	2	2	132	587.1	220.8
NC_022246.1	Streptococcus intermedius B196	4	4	304	586.0	198.7
NC_022236.1	Streptococcus constellatus subsp. pharyngis C232	4	4	267	588.4	213.5
NZ_CP021181.1	Sphingomonas sp. DC-6	2	2	478	558.8	153.8
NC_013521.1	Sanguibacter keddiei DSM 10542	4	4	302	583.3	208.9
NC_013520.1	Veillonella parvula DSM 2008	4	4	366	602.6	224.1
NC_018089.1	Streptococcus mutans GS-5	5	5	195	555.5	150.5
NZ_AP012334.1	Scardovia inopinata JCM 12537 DNA	2	2	107	590.4	287.9
NZ_CP018221.1	Sphingomonas sp. JJ-A5	2	2	348	562.5	178.6
NC_017768.1	Streptococcus mutans LJ23 DNA	5	5	203	572.1	172.1
NZ_CP019511.1	Sphingomonas sp. LM7	2	1	379	639.3	335.8
NC_020561.1	Sphingomonas sp. MM-1	2	2	475	594.1	254.4
NZ_CP009227.1	Treponema sp. OMZ 838	2	2	148	618.5	413.7
NC_007492.2	Pseudomonas fluorescens Pf0-1	6	6	484	573.8	178.3
NC_014034.1	Rhodobacter capsulatus SB 1003	4	4	722	650.8	328.0
NZ_CP019721.1	Veillonella parvula strain UTDB1-3	4	4	354	578.5	180.0

Таблица 4: Результаты анализа полноразмерных геномов (5'М домен)

NCBI ID	Name	Expected	Covered	FP-intervals	Length(avg.)	Length SD
NZ_CP012959.1	Aggregatibacter actinomycetemcomitans strain 624	6	6	35	954.4	175.2
NC_014640.1	Achromobacter xylosoxidans A8	3	0	41	976.6	232.5
NC_005966.1	Acinetobacter sp. ADP1 complete genome	7	7	14	974.8	207.5
NZ_CP009448.1	Achromobacter xylosoxidans C54	3	0	43	932.7	146.0
NC_013171.1	Anaerococcus prevotii DSM 20548	4	3	29	1174.6	609.3
NZ_CP012590.1	Actinomyces sp. oral taxon 414 strain F0588	3	3	28	1033.9	356.2
NZ_CP014060.1	Achromobacter xylosoxidans strain FDAARGOS_147	3	0	79	1024.9	343.2
NZ_CP007502.1	Aggregatibacter actinomycetemcomitans HK1651	6	6	17	884.0	64.0
NZ_LT635457.1	Actinomyces sp. Marseille-P2985	3	3	5	1120.8	217.2
NZ_CP012046.1	Achromobacter xylosoxidans strain MN001	3	0	56	953.9	206.3

Таблица 5: Результаты анализа полноразмерных геномов (5'М + центральный домены)

4 Заключение

Данный эксперимент показал наличие возможности использовать вторичную структуру цепочки для предварительной фильтрации кандидатов, однако требуется существенная доработка используемых алгоритмов. В дальнейшем планируется выполнение работ по следующим направлениям.

- Поиск, разработка и применение алгоритмов автоматического вывода грамматик для построения грамматики, описывающей вторичную структуру цепочки.
- Анализ ложных срабатываний и пропущенных кандидатов с целью выявить их особенности и доработать соответствующим образом алгоритм поиска.
- Разработка методов уменьшения количества ложных срабатываний.

Список литературы

- [1] Nawrocki E. P., Eddy S. R. Infernal 1.1: 100-fold faster RNA homology searches //Bioinformatics. — 2013. — Т. 29. — №. 22. — С. 2933-2935.
- [2] Dowell R. D., Eddy S. R. Evaluation of several lightweight stochastic context-free grammars for RNA secondary structure prediction //BMC bioinformatics. — 2004. — Т. 5. — №. 1. — С. 1.
- [3] Eddy S. R. A memory-efficient dynamic programming algorithm for optimal alignment of a sequence to an RNA secondary structure //BMC bioinformatics. — 2002. — Т. 3. — №. 1. — С. 1.
- [4] Yogev S., Milo N., Ziv-Ukelson M. StemSearch: RNA search tool based on stem identification and indexing //Bioinformatics and Biomedicine (BIBM), 2013 IEEE International Conference on. — IEEE, 2013. — С. 145-152.
- [5] Eddy S. R. Homology searches for structural RNAs: from proof of principle to practical use //RNA. — 2015. — Т. 21. — №. 4. — С. 605-607.
- [6] Anderson J. W. J. et al. Evolving stochastic context-free grammars for RNA secondary structure prediction //BMC bioinformatics. — 2012. — Т. 13. — №. 1. — С. 78.
- [7] Hemerik K. Towards a Taxonomy for ECFG and RRPg Parsing //Language and Automata Theory and Applications. — 2009. — С. 410-421.
- [8] Quast C. et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools //Nucleic acids research. — 2012. — Т. 41. — №. D1. — С. D590-D596.
- [9] YARD language description. Online. <http://yacconstructor.github.io/YaccConstructor/yard.html>

- [10] Rehakova K. et al. Variation in secondary structure of the 16S rRNA molecule in cyanobacteria with implications for phylogenetic analysis //Fottea. – 2014. – Т. 14. – С. 161-178.

Приложение

А Грамматика 16S на языке YARD, использовавшаяся в эксперименте

```
inline any: A | U | G | C
inline any_1_2: any*[1..2]
inline any_1_3: any*[1..3]
inline any_2_3: any any_1_2
inline any_2_4: any*[2..4]
inline any_3_4: any*[3..4]
inline any_3_5: any any_2_4
inline any_5_7: any any any_3_5
inline any_4_6: any any_3_5
inline any_6_8: any any_5_7
inline any_9_11: any*[9..11]
inline any_4 : any any any any
```

```
stem1<s>:
    A s U
  | U s A
  | C s G
  | G s C
  | G s U
  | U s G
  | A s G
  | G s A
```

```
stem2<s>: stem1<stem1<s>>
stem4<s>: stem2<stem2<s>>
stem6<s>: stem4<stem2<s>>
stem8<s>: stem4<stem4<s>>
```

```
mk_stem<s>:
    A mk_stem<s> U
  | U mk_stem<s> A
  | C mk_stem<s> G
  | G mk_stem<s> C
```

```

| G mk_stem<s> U
| U mk_stem<s> G
| G mk_stem<s> A
| A mk_stem<s> G
| s

```

```

stem<s>: mk_stem<stem4<s>>
stem_2<s>: mk_stem<stem2<s>>

```

```

stem_e1<s> : stem_2<(any stem_2<s> | stem_2<s> any)> | stem<s>
stem_e2<s> : stem_2<(any stem_e1<s> any | any stem_e1<s>
               | stem_e1<s> any)> | stem<s>
stem_4: stem_2<any_4>

```

```

[<Start>]
full: middle_part_root

```

```

head_part_root: h3
middle_part_root: h19
tail_part_root: h28 any_3_5 h44 any_3_5 h45

```

```

head_middle_folded: stem2<(any_6_8 h3 any_9_11 h19 any_1_2 h27 any_2_4)>
full_size_root: h3 any_9_11 h19 any_1_2 h27 any*[7..9] tail_part_root

```

```

(* 5'M domain *)

```

```

h3: stem_e2<(any_1_2 h4 any_1_3 h16 any_3_5
             (h17 | any*[1..6]) any*[2..5] h18 any_1_2)>
h4: stem_e1<(h5 h15 any?)>
h5: any_5_7 stem_e2<(any_1_3 h6 any_5_7
                    stem_2<(any_5_7 h7 any? h11 any_1_3 h12 any?)>
                    any_1_2 h13 any_1_2 h14 any_2_4)> any_3_5

```

```

h6: stem_e2<stem_e2<stem_e2<stem_e2<any_3_4>>>>
h7: stem_e2<(any_2_4 stem<(any_1_2 h8 any_4_6 h9 any_3_5 h10 any_1_2)>
             any_1_3)>
h8: stem_2<(any_3_5 stem_4 any_3_5)>
h9: stem_2<any_3_5>
h10: stem_e2<any_3_5>
h11: stem_2<(any_2_4 stem_e2<any_6_8> any_3_5)>
h12: stem<(any? stem_2<any_3_5> any_2_4)>
h13: stem<any_9_11>
h14: stem_2<any_3_5>

```

```

h15: stem_e1<(any_2_4 stem2<any_4> any?)>
h16: stem_2<(any_5_7 stem_2<any_2_4> any_4_6)>

```

```

h17: stem<(any*[6..9] stem_2<any*[7..11]> any_6_8)>
h18: stem<(any_5_7 stem<(any_4_6 stem_2<any_3_5> any_6_8)>>>

(* Central domain *)
h19: stem_2<(any_5_7 h20 any_3_5 h25 any*[9..12] h26 any_1_2)>
h20: stem_2<( any_3_4 stem_2<( any_1_2 h21 any_2_4 h22 any_2_4 )> any_3_4 )>
h21: stem_e2<( any_3_5 stem_e2<(any_3_5 stem_e1<any*[5..6]> any_2_4)> any_3_5 )>
h22: stem_e2<( any_1_3 stem<(any_3_4 h23 any*[10..12] stem_2<( any any A any )>
    any_1_2)> any_1_3 )>
h23: stem<(any_2_4 stem_2<any*[5..6]> any_5_7)>
h25: stem<(any*[7..11] stem<any*[8..10]> any*[4..7])>
h26: stem_e1<(any_1_2 stem_e2<any_4_6> any_3_5 stem_4 any_3_5 )>
h27: stem_2<(any_5_7 stem_4 any_3_5)>

(* 3'M domain *)
h28: stem_e2<(any h28_a any_2_4)>
h28_a: stem<(any_1_3 h29 any_4_6 h43 any_4_6)>
h29: stem<(h30 any_2_4 h41 any_5_7 h42 any_4_6)>
h30: stem_e1<(any_3_5 h31 any*[7..9] h32 any_2_4)>
h31: stem<any*[7..9]>
h32: stem<(any_4_6 h33 any_1_2 h34 any_3_5)>
h33: stem<(any_1_3 stem<any_4> any_1_3 stem<any_4> any_1_3)>
h34: stem_e1<(any_1_2 stem<(stem_e2<(any_2_4 h35
    any_4_6 h38 any_3_5)> any_2_4)>>
h35: stem<(h36 any_2_3 h37 any_2_3)>
h36: stem<any_4>
h37: stem<any_5_7>
h38: stem<(any_1_2 h39 any_1_3 h40 any_4_6)>
h39: stem<(any_2_4 stem<(any_1_3 stem<any_4_6>)> any_2_4)>
h40: stem<any_4>
h41: stem<(any_4_6 stem<(any_1_3 stem<(any_2_4 stem<any_4> any_2_4)>
    any_3_5)> any_4_6)>
h42: stem<(any_3_4 stem<any*[7..9]> any_3_4)>
h43: stem<any*[7..9]>

(* 3'm domain *)
h44: stem<(any_1_3 stem<(any_2_4 stem<(any_1_3 stem<(any_3_5
    stem_e1<(any_1_3 stem<any_4>)> any_2_4)> any_1_3)> any_3_5)> any_2_3)>
h45: stem<any_4>

```