План тестирования алгоритмов Haussler’а и Eades’а.

Изначально собирается граф со следующими параметрами.

Это разумная редукция статистических данных на случай когда в графе 8’000, а не 3’234’830’000 (полный геном человека) вершин.

|  |  |  |
| --- | --- | --- |
|  | Graph | |
| count | size |
| Bases in backbone | 8000 | 1 |
| Mobile elements | 2x6 | 30 |
| Large deletions | 8 | 45-500 |
| Duplications | 3 | 3-1000 |
| Short insertions | 30 | 2-20 |
| Short deletions | 30 | 2-20 |
| SNP | 300 | 1 |
| Total bases (aprox.) | 8700 |  |

Для него применяем сортировки Haussler’а и Eades’а и сравниваем результаты по двум метрикам: числу обратных ребер (reversed edges) и средней ширине разреза (average cut width).

Далее составляем графы с параметрами:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | Graph №1 | Graph №2 | Graph №3 | Graph №4 | Graph №5 | Graph №6 | size |
| count | count | count | count | count | count |
| Bases in backbone | 8000 | 8000 | 8000 | 8000 | 8000 | 8000 | 1 |
| Mobile elements | 2x6 | 4x6 | 6x6 | 8x6 | 10x6 | 12x6 | 30 |
| Large deletions | 8 | 16 | 24 | 32 | 40 | 48 | 45-500 |
| Duplications | 3 | 6 | 9 | 12 | 15 | 18 | 3-1000 |
| Short insertions | 30 | 60 | 90 | 120 | 150 | 180 | 2-20 |
| Short deletions | 30 | 60 | 90 | 120 | 150 | 180 | 2-20 |
| SNP | 300 | 600 | 900 | 1200 | 1500 | 1800 | 1 |

Это соответствует увеличению числа особей (число всех типов вариаций одновременно пропорционально растет, размер фрагмента не меняется). При необходимости можно продолжить рост.

Для каждого набора параметров (от 1 до 6) генерируем 100 (или больше?) графов, применяем ко всем к ним сортировки Haussler’а и Eades’а, сравниваем результаты (средний, лучший, худший).

Далее экспоненциально (х2, х4, х8, х16, х32) увеличиваем число вариаций только одного из типов. Тестируем по тому же принципу что и выше.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Increasing number of Mobile elements** | Graph №1 | Graph №2 | Graph №3 | Graph №4 | Graph №5 | Graph №6 | size |
| count | count | count | count | count | count |
| Bases in backbone | 8000 | 8000 | 8000 | 8000 | 8000 | 8000 | 1 |
| Mobile elements | **2x6** | **4x6** | **8x6** | **16x6** | **32x6** | **64x6** | 30 |
| Large deletions | 8 | 8 | 8 | 8 | 8 | 8 | 45-500 |
| Duplications | 3 | 3 | 3 | 3 | 3 | 3 | 3-1000 |
| Short insertions | 30 | 30 | 30 | 30 | 30 | 30 | 2-20 |
| Short deletions | 30 | 30 | 30 | 30 | 30 | 30 | 2-20 |
| SNP | 300 | 300 | 300 | 300 | 300 | 300 | 1 |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Increasing number of Large deletions** | Graph №1 | Graph №2 | Graph №3 | Graph №4 | Graph №5 | Graph №6 | size |
| count | count | count | count | count | count |
| Bases in backbone | 8000 | 8000 | 8000 | 8000 | 8000 | 8000 | 1 |
| Mobile elements | 2x6 | 2x6 | 2x6 | 2x6 | 2x6 | 2x6 | 30 |
| Large deletions | **8** | **16** | **32** | **64** | **128** | **256** | 45-500 |
| Duplications | 3 | 3 | 3 | 3 | 3 | 3 | 3-1000 |
| Short insertions | 30 | 30 | 30 | 30 | 30 | 30 | 2-20 |
| Short deletions | 30 | 30 | 30 | 30 | 30 | 30 | 2-20 |
| SNP | 300 | 300 | 300 | 300 | 300 | 300 | 1 |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Increasing number of Duplications** | Graph №1 | Graph №2 | Graph №3 | Graph №4 | Graph №5 | Graph №6 | size |
| count | count | count | count | count | count |
| Bases in backbone | 8000 | 8000 | 8000 | 8000 | 8000 | 8000 | 1 |
| Mobile elements | 2x6 | 2x6 | 2x6 | 2x6 | 2x6 | 2x6 | 30 |
| Large deletions | 8 | 8 | 8 | 8 | 8 | 8 | 45-500 |
| Duplications | **3** | **6** | **12** | **24** | **48** | **96** | 3-1000 |
| Short insertions | 30 | 30 | 30 | 30 | 30 | 30 | 2-20 |
| Short deletions | 30 | 30 | 30 | 30 | 30 | 30 | 2-20 |
| SNP | 300 | 300 | 300 | 300 | 300 | 300 | 1 |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Increasing number of Short InDels** | Graph №1 | Graph №2 | Graph №3 | Graph №4 | Graph №5 | Graph №6 | size |
| count | count | count | count | count | count |
| Bases in backbone | 8000 | 8000 | 8000 | 8000 | 8000 | 8000 | 1 |
| Mobile elements | 2x6 | 2x6 | 2x6 | 2x6 | 2x6 | 2x6 | 30 |
| Large deletions | 8 | 8 | 8 | 8 | 8 | 8 | 45-500 |
| Duplications | 3 | 3 | 3 | 3 | 3 | 3 | 3-1000 |
| Short insertions | **30** | **60** | **120** | **240** | **480** | **960** | 2-20 |
| Short deletions | **30** | **60** | **120** | **240** | **480** | **960** | 2-20 |
| SNP | 300 | 300 | 300 | 300 | 300 | 300 | 1 |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Increasing number of SNPs** | Graph №1 | Graph №2 | Graph №3 | Graph №4 | Graph №5 | Graph №6 | size |
| count | count | count | count | count | count |
| Bases in backbone | 8000 | 8000 | 8000 | 8000 | 8000 | 8000 | 1 |
| Mobile elements | 2x6 | 2x6 | 2x6 | 2x6 | 2x6 | 2x6 | 30 |
| Large deletions | 8 | 8 | 8 | 8 | 8 | 8 | 45-500 |
| Duplications | 3 | 3 | 3 | 3 | 3 | 3 | 3-1000 |
| Short insertions | 30 | 30 | 30 | 30 | 30 | 30 | 2-20 |
| Short deletions | 30 | 30 | 30 | 30 | 30 | 30 | 2-20 |
| SNP | **300** | **600** | **1200** | **2400** | **4800** | **9600** | 1 |

Наконец, для тестирования времени работы, мы увеличиваем число вершин в backbone и параллельно увеличиваем число вариаций:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | count | count | count | count | count | count |
| Bases in backbone | 8000 | 16000 | 32000 | 64000 | 128000 | 256000 |
| Mobile elements | 2x6 | 4x6 | 8x6 | 16x6 | 32x6 | 64x6 |
| Large deletions | 8 | 16 | 32 | 64 | 128 | 256 |
| Duplications | 3 | 6 | 12 | 24 | 48 | 96 |
| Short insertions | 30 | 60 | 120 | 240 | 480 | 960 |
| Short deletions | 30 | 60 | 120 | 240 | 480 | 960 |
| SNP | 300 | 600 | 1200 | 2400 | 4800 | 9600 |