

MiSeqでの*de novo*実験

ILLUMINA Webinar

OIST / 20130412 16:00-17:00

DNA Sequencing Section (SQC)
小柳 亮、藤江 学



OIST

OKINAWA INSTITUTE OF SCIENCE AND TECHNOLOGY GRADUATE UNIVERSITY

本日のトピックス

Wet 体が疲れたら汗を拭いて
ドライ！

Dry 目が疲れたら汗をかけ！

基本は水陸両用！

本日の担当

Wet

藤江

Dry

小柳

実験フロー

DNA
抽出

断片化

精製

DNA
濃度
測定

Size
測定

EndRepair

dA-
tailing

Ligation

PCR
増幅

Size
Select

Library
定量

実験を始める前に

サンプルの確認

- 何倍体？
- 野生？近交系？
- ゲノムサイズは？
- サンプル量は？

DNA抽出

長いゲノムDNAはなぜ必要か

- *de novo*解析に必須なMatePair解析をするため

PCI処理とRNase処理の重要性

- PCI処理にはPhase Lock Gelが最適
- RNase未処理のDNA濃度 → DNA+RNA濃度

断片化直前の濃度測定がなぜ必要か

- ゲノムDNAは溶解しにくく、濃度が変化しやすい

DNA断片化

酵素と超音波断片化の違い

- ・ DNA純度の影響 酵素 > 超音波
- ・ 不純物の多いDNAは超音波の方が断片化サイズが安定

Covaris社製チューブ

- ・ MicroTUBE < 2Kbp
- ・ 2Kbp < miniTUBE < 5Kbp
- ・ 5Kbp < g-TUBE < 30Kbp

DNA精製

なぜ断片化後に行うのか

- 長いDNAの精製方法が限られている
- 長いDNAは精製によりダメージを受けやすい
- 短いDNAは精製しやすく切れにくい

AMPureとQiaQuick

- 精製によって除去されるものが異なる

DNA精製

サンプル DNA	260/280	260/230
未精製	1.79	1.02
AMPure	1.69	0.89
AMPure + QIAQuick	1.75	2.47

DNA濃度測定

吸光測定と蛍光測定は何が違うのか

- 正確性は 蛍光測定 > 吸光測定
- DNAの精製度が高ければ蛍光測定 \div 吸光測定

どの段階のDNA量を信頼すればよいのか

- 「断片化→精製後のDNA量」を実験の指標にする

DNA Size測定

DNA断片のサイズ情報はなぜ必要か

- クラスター形成効率に影響する
- DNA重量よりも分子数が重要

分子数は 5ug 300bp > 5ug 500bp

酵素反応

EndRepair

dA-tailing

- A付加後に実験を中断しない

Ligation

- テンプレートとアダプターの分子数比

PCR増幅

PCRを行なってデータが良くなることがあるか

- それはない
- *de novo*解析はPCR-Freeが良い

PCR前のライブラリコピー数の重要性

PCR増幅

	ライブラリの コピー数 (PCR増幅後)
ライブラリA	2.10E+08
ライブラリB	1.64E+08

	ライブラリの コピー数 (PCR増幅前)	PCRサイクル数	ライブラリの コピー数 (PCR増幅後)
ライブラリA	200	20	2.10E+08
ライブラリB	160000	10	1.64E+08

Library Size Selection

なぜサイズセレクトを行うのか

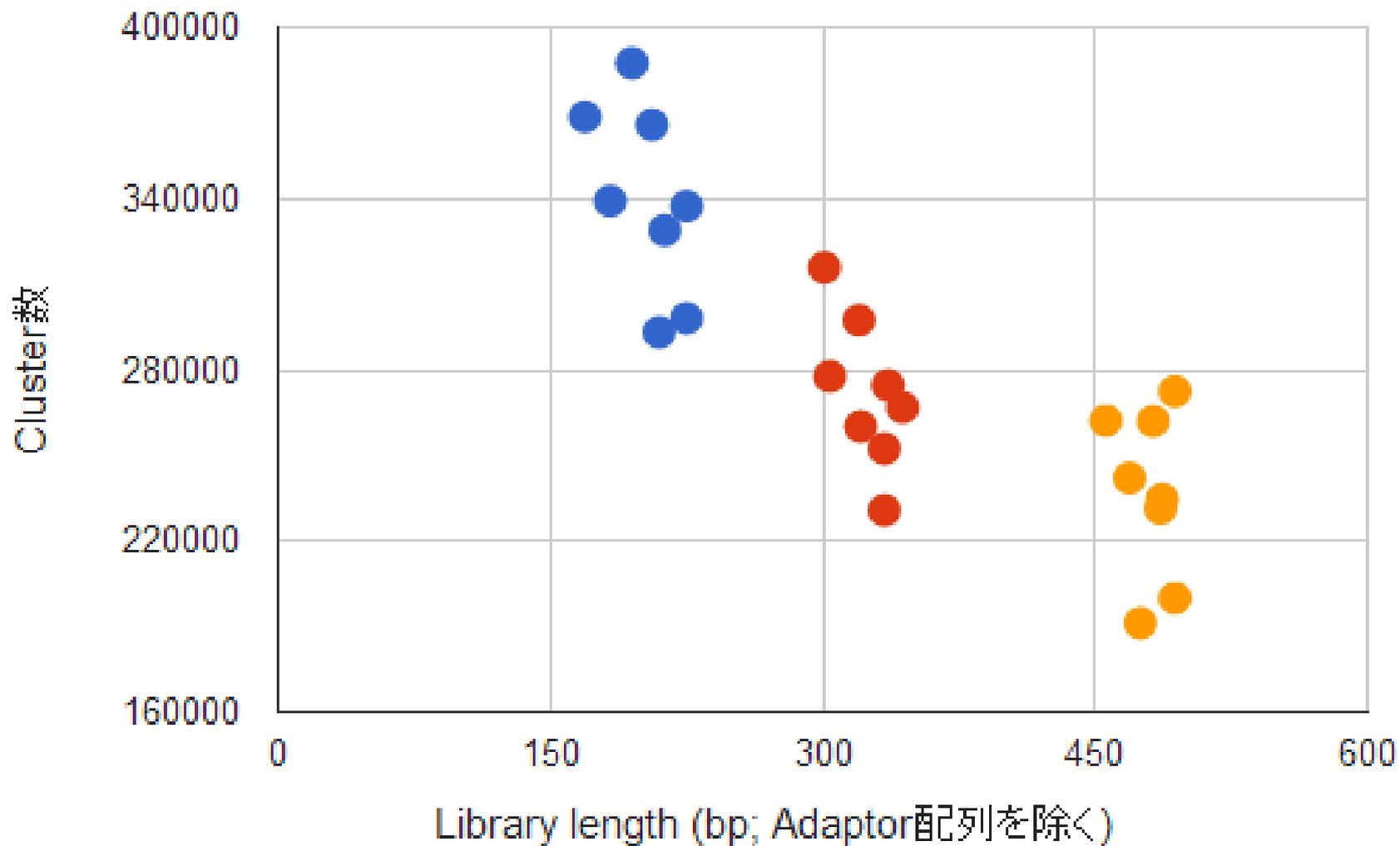
- リアルタイムPCRなどの定量を可能にするため
- クラスター数を安定させるため
- アセンブル時のパラメーターで必要になるため

PippinPrepとE-gelとAMPure

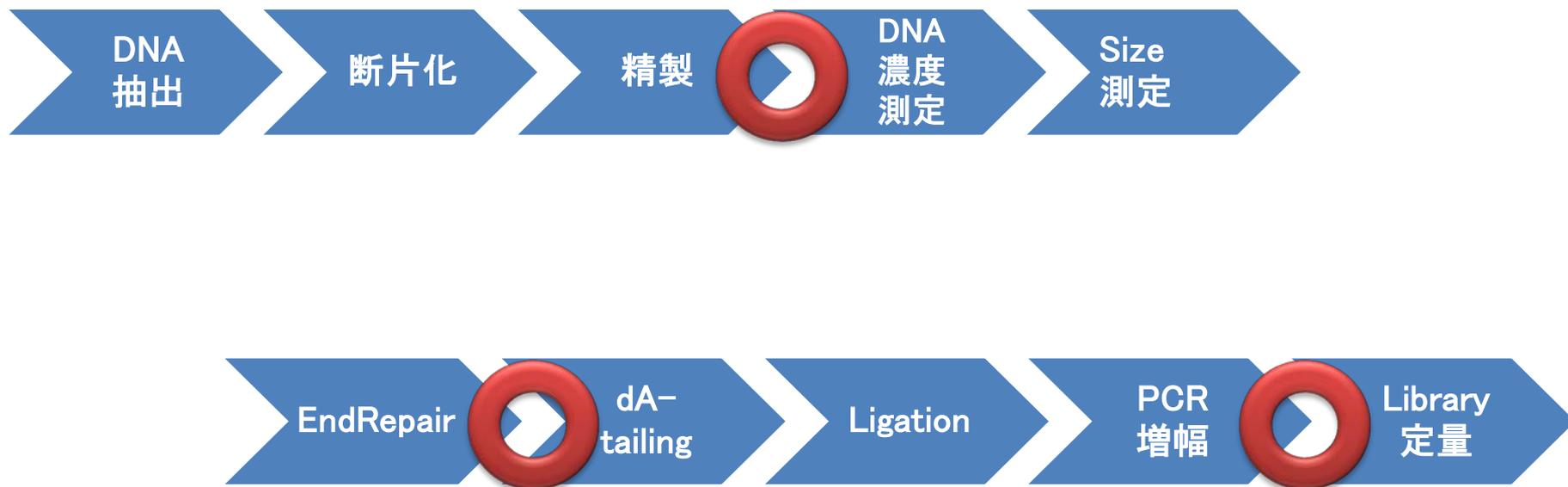
Library Size Selection

方法	精度	処理 スピード
BluePippin	◎	△
E-Gel	○	△
AMPure	△	◎

Library Size Selection



Library Size Selection



Library定量

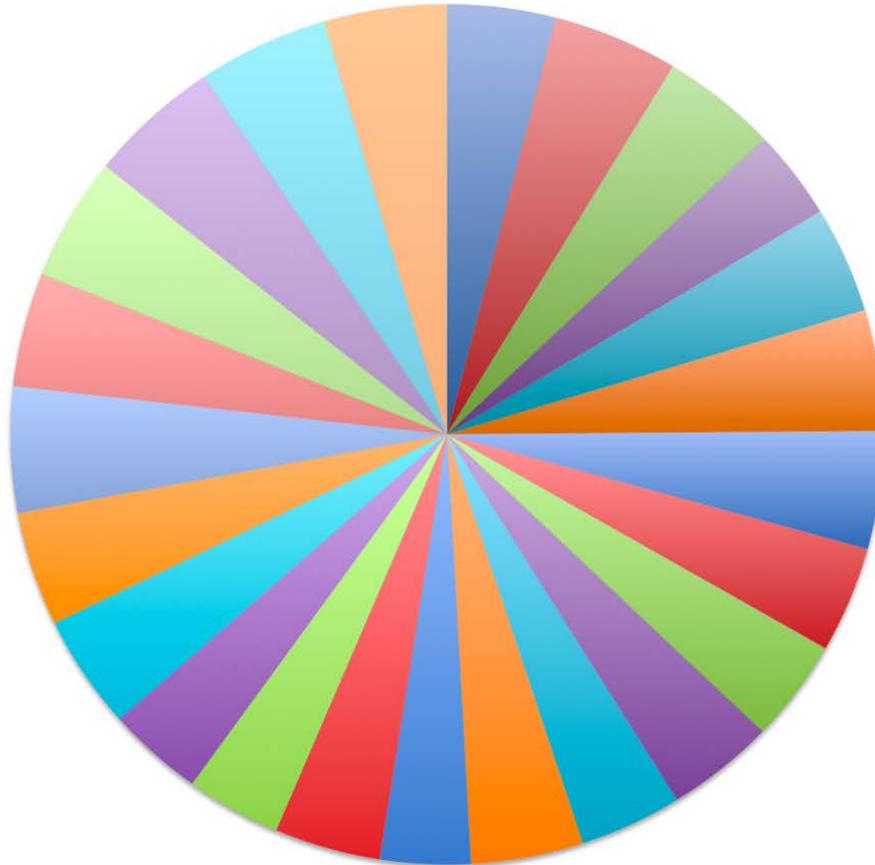
なぜLibrary定量を行うのか

- クラスタ数安定させるため
- Multiplex解析で均等なデータ量を得るため

電気泳動と相対定量と絶対定量

Library定量

Yield (Mbases)



Digital PCR (Bio-Rad)

Library定量

方法	精度	处理能力
Bioanalyzer	△	○
Quantitative PCR	○	△
Digital PCR	◎	◎

Adapterの作り方

配列はIllumina社に問い合わせ

全てのPrimerの両末端2塩基はS化处理

Indexが含まれるプライマーの5末端はリン酸化

Oligoを溶かすバッファー(TEに塩少々)

- 10mM Tris-HCl pH8.0
- 1mM EDTA
- 50mM NaCl

Adapter Annealing Program

- 95°C 10min
- 95°C 4sec
- 0.1°C 4secずつ温度低下で15°Cまで下げる

長いDNAの扱い方

精製方法

- Purification by Genomic DNA Clean & Concentrator (Zymo reserch)

切出方法

- Thermostable β -Agarase (ニッポンジーン)

長いDNAの使い方

方法	回収量	三回精製時の回収量 (%)
QIAEX II Gel Extraction Kit	20~40%	0.8~6.4%
Genomic DNA Clean & Concentrator + Thermostable β -Agarase	70~90%	34.3~72.9%

Miseqの扱い方

簡単すぎて書くことがありません



次はDry (解析編) です

De novoゲノムシーケンシングのワークフロー



シーケンスデータの比較

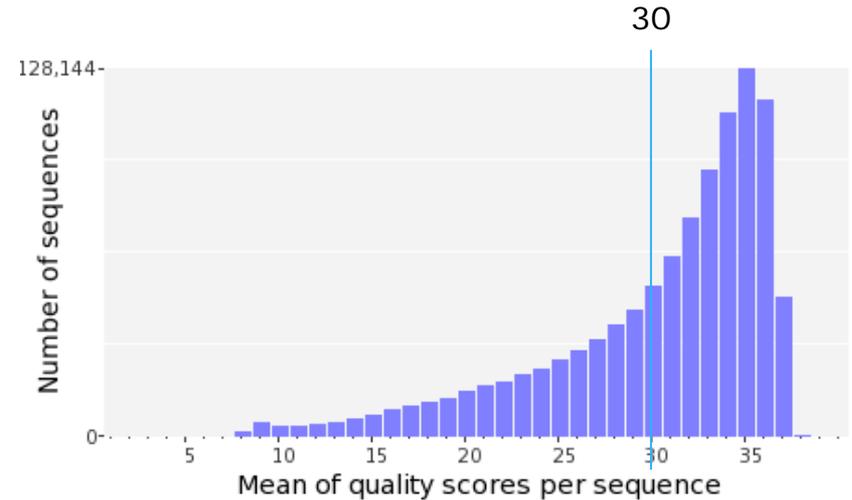
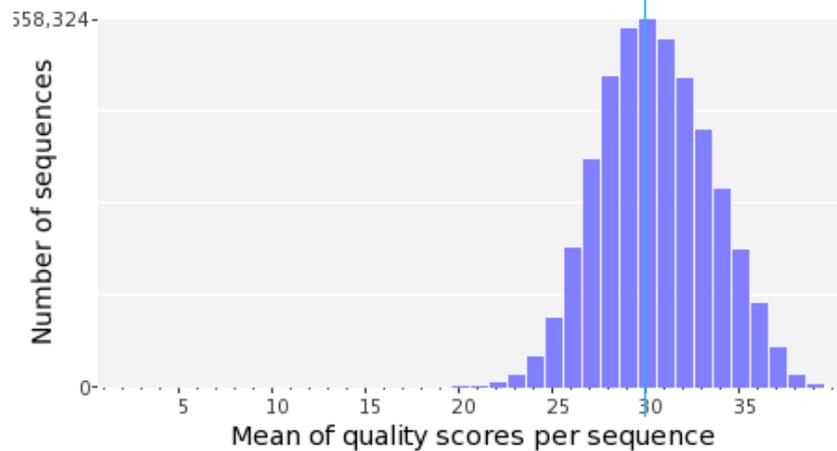
454 FLX+

Illumina MiSeq (250bp)

Quality Score at Read Position

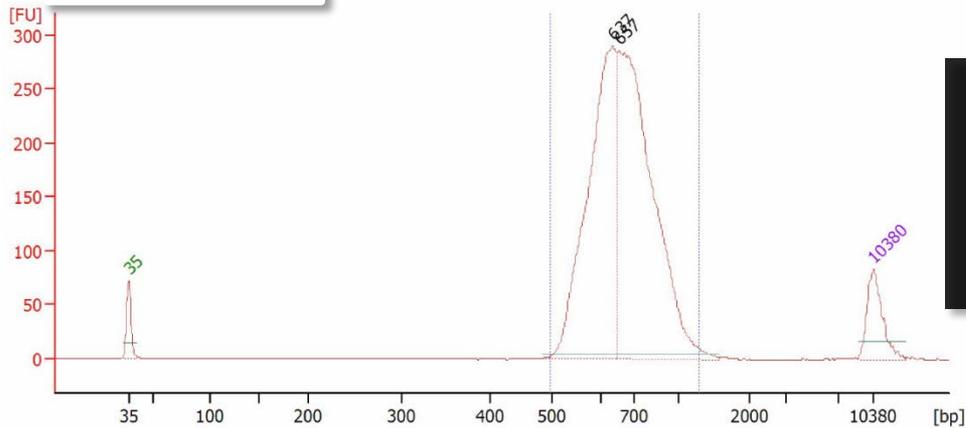


Mean Quality Scores per Sequence



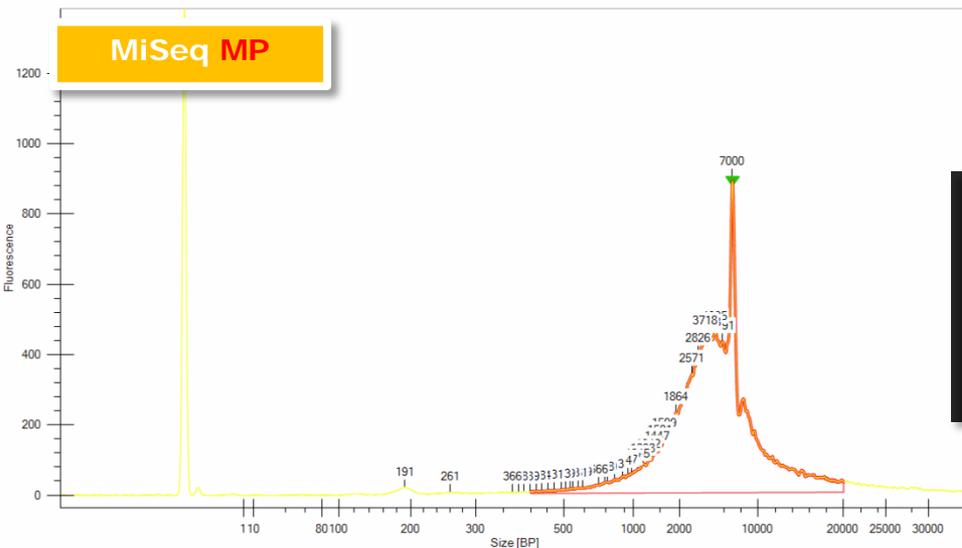
Paired-End, Mate-Pairリードのマッピング状況

MiSeq PE



```
library
{
  pairDistanceRangeUsed      = 314..1004;
  computedPairDistanceAvg    = 628.8;
  computedPairDistanceDev    = 187.8;
}
```

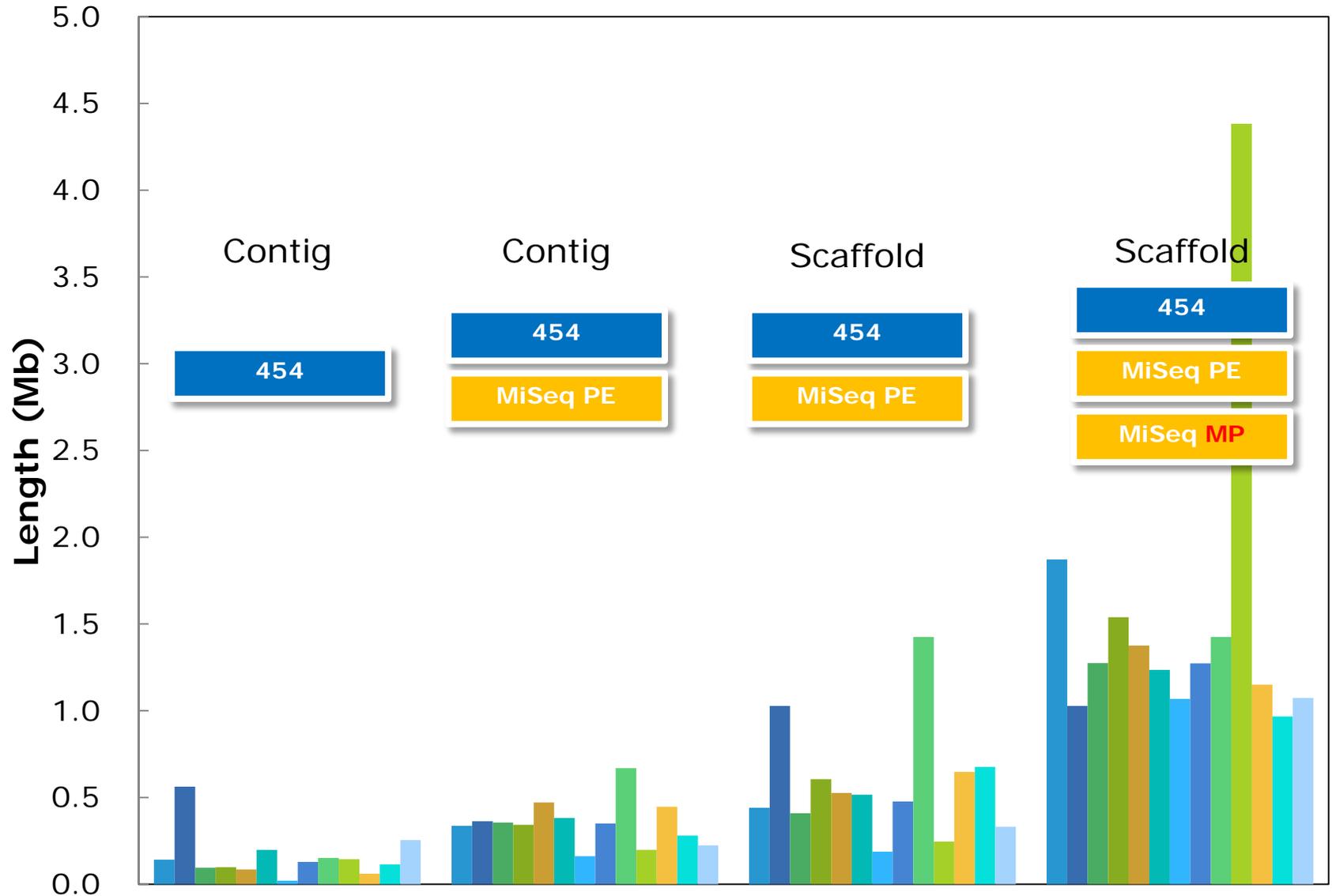
MiSeq MP



```
library
{
  pairDistanceRangeUsed      = 1251..4110;
  computedPairDistanceAvg    = 2502.3;
  computedPairDistanceDev    = 804.1;
}
```

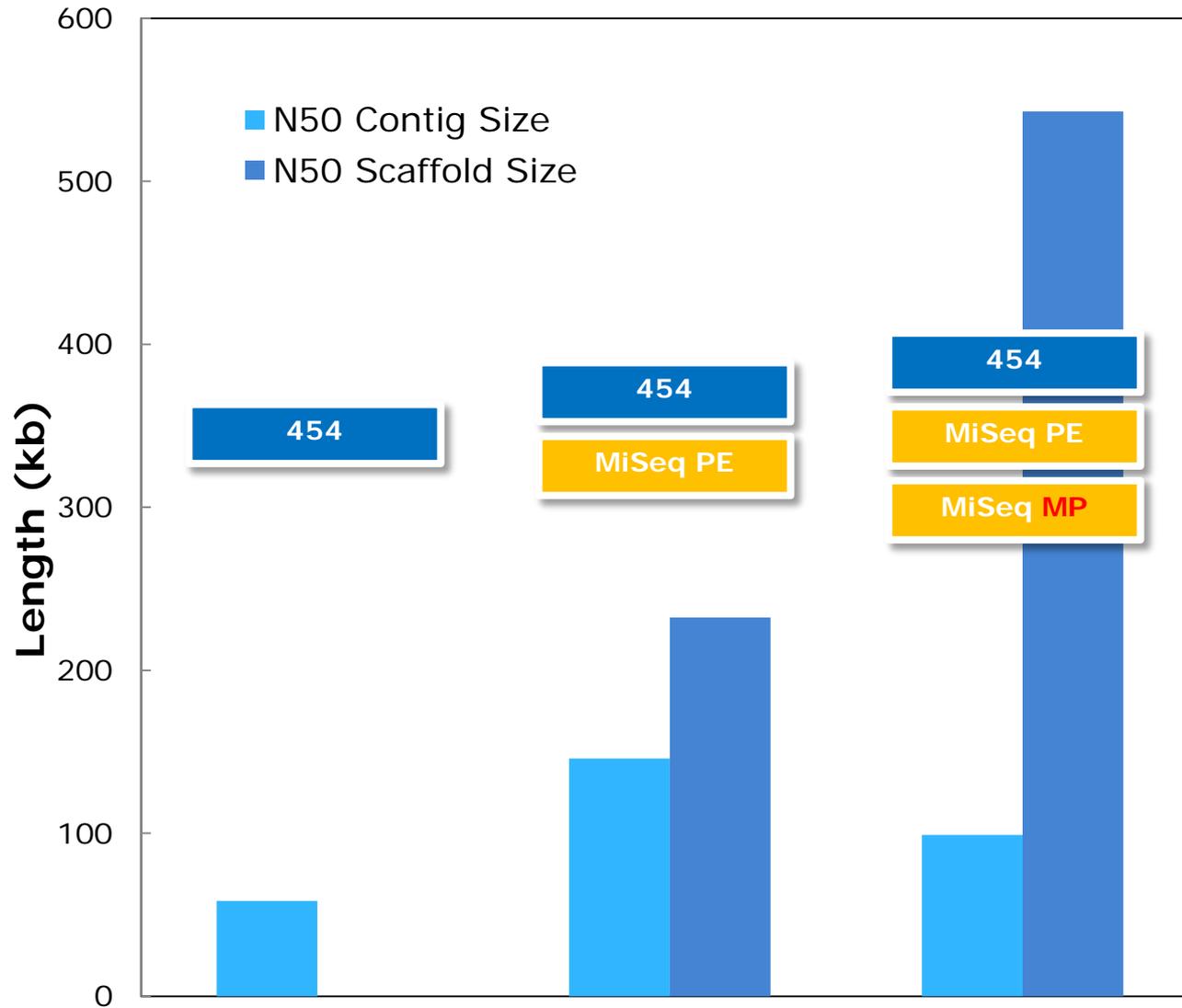
微生物ゲノムのアセンブル 1

放線菌13株 予想ゲノムサイズ : 4.1-16 Mb



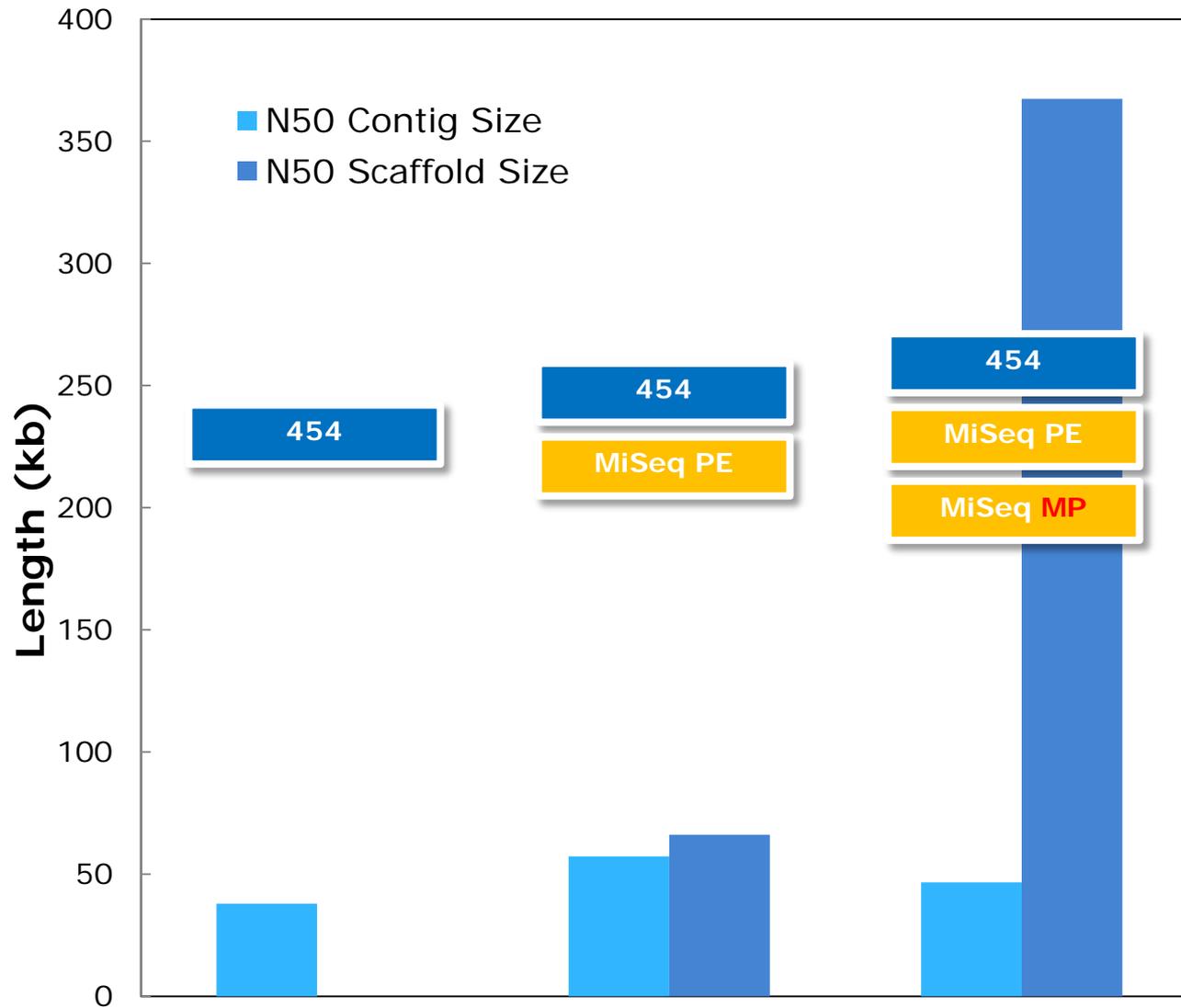
微生物ゲノムのアセンブル 2

Aurantiochytrium sp. 予想ゲノムサイズ : **40 Mb**



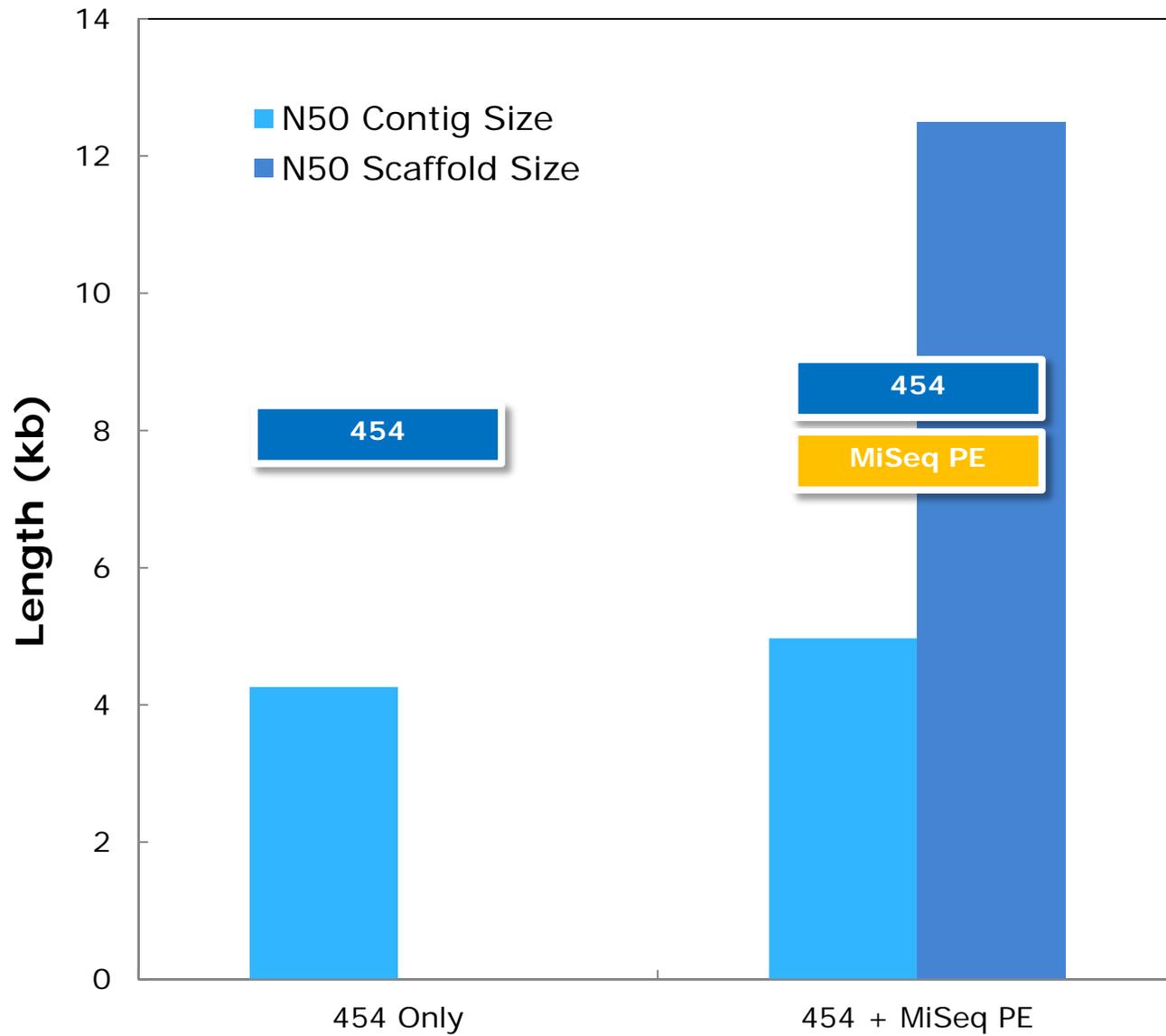
微生物ゲノムのアセンブル 3

Aurantiochytrium sp. 予想ゲノムサイズ : **80 Mb**



無脊椎動物ゲノムのアセンブル

L. anatina 予想ゲノムサイズ : **400 Mb**



De novoゲノムシーケンシングのワークフローは変わるのか？



シーケンス後のデータ解析環境について

- **ハードウェア**

- ゲノムサイズが100Mbくらいまでなら市販のパソコンでもある程度対応可能
 - ※ただし上位機種に限る
- それ以上はより多くのメモリが必要になる

- **ソフトウェア**

- OSはLinux
 - ソフトの選択肢が最も広い
 - デイトリビューションは使いたいソフトの要求する環境に合わせる

- **OIST SQCでは**

- **個人用**
 - Core i7, 48GB RAM, Fedora/Ubuntu
- **SQC共用**
 - 4x Xeon, 512GB RAM, Ubuntu/Scientific Linux 6
- **OIST HPC**
 - Large Memory Nodes: 1TB RAM, Scientific Linux 5
 - Cluster: 192 nodes, Scientific Linux 5/6

自作自作

- **MiSeq単独なら**
 - Velvet
 - ABySS
 - ALLPATH-LG
 - MIRA
- **454+MiSeqなら**
 - Newbler
 - Celera Assembler

それぞれ得意・不得意があるので試行錯誤が必要

- **不必要な配列の除去**
 - 低クオリティ
 - 低複雑度
 - アダプタ配列
- **ソフトの要求する形式への変換**
 - 特にPaired-End, Mate-Pairの場合。
 - \leftrightarrow 、 $\rightarrow\leftarrow$ 、 $\rightarrow\rightarrow$
 - 単一ファイルか複数ファイルか
 - 多くの変換は既存のツールでできる
 - ※ただしコマンド入力が必要
- **ソフトにより仕様が異なるので、説明書をよく読む**
 - 説明書だけでなくメーリングリストも読む

- **SEQanswers**
 - <http://seqanswers.com/>
- **NGS現場の会**
 - <http://ngs-field.org/>