Scalable Multiplexed Library Preparation for Low-Depth Sequencing and Genotype Imputation

Ariele Hanek¹, Christopher Fragoso², Rebecca Feeley¹, Jack Leonard¹, Joseph Mellor¹

(1) seqWell, Inc. 66 Cherry Hill Drive, Beverly MA 01915 USA (2) Verinomics, Inc. 5 Science Park, New Haven CT 06511 USA

Summary

The use of low-pass whole-genome sequencing (WGS) is a robost and cost-effective alternative approach to microarray-based genotyping for genotype imputation pipelines. With prior knowledge of known parental or population haplotypes, genotypes can be accurately imputed from 'shallow' genomic sequencing of individual plant or animal samples at raw coverage depths as low as 0.5x.

The economics of using sequencing in place of microarrays depends on the ability to reliably create balanced pools of libraries that efficiently multiplex as many genomic samples as possible on large high-output sequencing runs that achieve the lowest cost per unit of data.

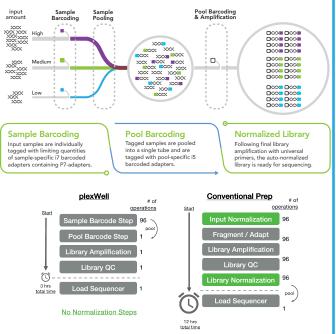
plexWell[™] library preparation technology streamlines the ability to create normalized multiplexed libraries from multiple samples in batches of 96 to 384 samples without the need for time-consuming measurement or adjustment of input DNA concentrations, significantly simplifying the complex task of high-level multiplexing.

Methods

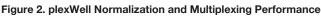
plexWell Library Preparation Technology

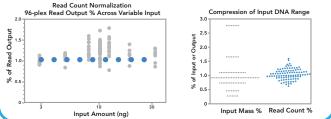
The technical foundation of the plexWell approach (Fig 1) is a reagent-limited initial transposition step performed on many samples in parallel, coupled to a subsequent pooled library generation step; when applied in conjunction, these two steps yield an approximately equal number of sequencer-ready library fragments from each of a potentially large collection of samples across a 10-fold input range.





The plexWell[™] library workflow (above) reduces the time and effort associated with multiplexed library prep, while maintaining multiplex uniformity across variable input DNA concentrations (Fig 2, below)





Results

139 individuals of a rice recombinant inbred line (RIL) population, derived from an *indica* x tropical *japonica* F₁ hybrid and advanced through single seed descent for 8 generations, were reanalyzed through a workflow of Verinomics DNA extraction, seqWell library preparation, and whole genome sequencing via Illumina NovaSeq 6000 2x150. Sequencing read alignment, variant calling, variant filtering, and imputation were performed by the Verinomics population genomics pipeline. The imputed WGS dataset contained an average of 940,897 SNPs per sample (Table 1., Fig. 3A) and were so et al. 2017). In addition to being highly accurate, the imputed WGS genotypes facilitate gene-level functional analysis for trait dissection (Fig. 4).

Table 1. Summary Statistics for N = 139 Rice RIL Sequencing

| | Total mapped reads | Total genomic coverage (MB) | Coverage per base | Raw SNPs | Filtered SNPs | Imputed SNPs | Fraction of missing imputed calls | | |
|---------|--------------------------|--------------------------------------|----------------------|-------------|------------------|-----------------|--|--|--|
| Average | 9,510,767 | 1,427 | 3.82 | 1,360,634 | 572,211 | 940,897 | 0.08 | | |
| SD | 4,146,824 | 622 | 1.67 | 526,448 | 224,367 | 123,089 | 0.03 | | |

Figure 3. (A) Histograms of 139 Rice RIL Summary Statistics of Coverage and Imputed Genotype Calls and (B) Concordance with Genotyping by Sequencing

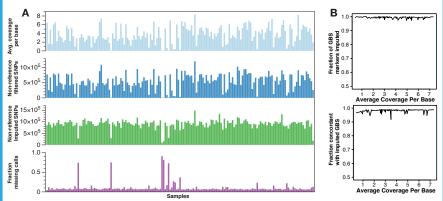


Figure 4. Low-Depth WGS Facilitates Gene-Level Functional Analysis

| Converse File Converse File Converse Co | 4,000,000 6,000 | | 13,30 quence | 0 12,000,000 12,500 | 14,000,000 | 2 0 | 000 | chrtti - chrt 18.000.000 Zoorri in to | 210,0 | 00,000 | 14028 (16.3 22,000, 13,307,590 | ,000 | 24,993,9 | see seque | 13,319 Hote | 2 5580-1. | oom in t | 00 Share 000,000 2 0 See Sequence | Falson 2,000,000 13,312,5 | 24. | 90,093 | se,o |
|---|--|--|--|---|---|---|---|---|--|---|---|---|--|--|---|---|--|--|--|--|---|---|
| Parferences, yequences 1955P-10_represent 1922_517.pH 1923_517.pH 1933_514.qH 1933_514.qH | 12,330,000 | Zoom in to see se | 19,30 rojuence Georgeon III III III | 2,599 Zoom k | 14,000,000 | 16,000 | 000 | 18,003,000 Zoom in to | 210,0 | 00,000 | 22,000, | ,000 | 24,993,9 | see sequer | 12,31 | 2,000 20 05400 1. | oom in t | o see sequence | 13,312,5 | 34, | | |
| Parferences, yequences 1955P-10_represent 1922_517.pH 1923_517.pH 1933_514.qH 1933_514.qH | 12,330,000 | Zoom in to see se | 19,30 rojuence Georgeon III III III | 2,599 Zoom k | n to see sequenc | 13,8 | (66) | Zoom in to | | | | 2 | oom in to | see seque | 12,31 | 2,000 20 05400 1. | oom in t | o see sequence | 13,312,5 | 34, | | |
| 18059-1.0_represent 1927_5187.gH 1929_838.gH 1931_594.gH 1935_5185.gH | | | | | | 1 | | | see seq | uence | 1 | z | _ | | | 9580-1. | | | | | Zoom | n to see |
| 1927_5117.gH 1929_533.gH 1935_54.gH 1935_54.gH | otive_transcripts.gtf5 | | 10 10 10 | steen antipoten tran | emeson 1. | 110 | | | I | | Į | | 09030 | 353701-01 | | | | | | | | |
| 1927_5117.gH 1929_533.gH 1935_54.gH 1935_54.gH | obve_transcripts.pHS | | 10 10 10 | SSEC0-01 | emission t. | 110 | | _ | l | | Į | | 040300 | 1 | oolon bansmi | | | ciii - | | | | - |
| 1929_333.4/r 1931_594.4/r 1933_595.4/r | | | - 11 | 1 | 1 | 110 | | | I. | 1 | 1 | | 1 | 1 | 1 | | | CHI - | | | | |
| 1931_594.gft 1953_595.gft | | | - 11 | | i. | 110 | | 1 | ÷. | | | | | | | | | | | | | |
| 1931_594.gft 1953_595.gft | | | ш | | i. | 110 | | | | | | | | | | | | | | 11 | | |
| 1955 S165 gr | | | | | | | | | 1.1 | | | | | | | 1101 | | | | | | |
| | 11 | 1 | 11 | | | 1.1.1 | | | 1 | | | 11 | | | | 1111 | 1 1 | | | | | 1.1 |
| 1934_92.011 | | | | | 1 | 11 | | 1 | 1 | 1 | 1 | 11 | 1 | 1 | 1 | 1001 | 1 1 | 00 | | | | 1.1 |
| | | | 111 | 1.1 | 1 | 11 | | 1 | | | 11 | 11 | 1 | 1 | 1.1 | 1000 | t II | cini – | | 11 | | 1.1 |
| 1935 513.orf | | | | 1.1 | 1 | 111 | | | | | | 111 | 11 | 1 | 1.1 | 100 | t in | cine – | | | | 1.1 |
| | | | | | | 110 | | | ÷. | | 11 | 111 | - 0 | | | | 10 | | | | | 0.0 |
| 1936_323.gfr | 11 | | 10 | | | | | | 1 | | | | | | | | | | | | | |
| 1937_534.gff | 11 | | 10 | 1 | 1 | 11 | | 1 | 1 | | | 11 | 1 | | | 1101 | 1 1 | | | | | 11 |
| 1956_583.gH | 11 | 11 | 11 | 1.1 | 1 | 11 | | 1 | 1 | 1 | | 11 | 11 | 1 | 1.1 | 1001 | 1.1 | | | 1 | | 1.1 |
| 1939 995.ctf | 11 | | 10 | 1.1 | 1 | 11 | | 1 | 1 | | | 11 | 1.1 | 1 | | 1001 | t i | | | | | 1.1 |
| | | | | | | | | | <u> </u> | | - 1 | 111 | - 6 | | - 1 | | | | | | | 11 |
| | | | | | | 110 | | | 11 | | | | 11 | | | | | | | | | 11 |
| 1943_\$107.011 | 11 | | 111 | | 1.1 | | | | 1 | | | 11 | | | | | 1 1 | | | | | 11 |
| 1243_514.gtr | 11 | | 111 | 1.1 | 1 | 11 | | 1 | 1 | | | 11 | 1 | 1 | 1.1 | 1101 | 1.1 | 00 | | | | 1.1 |
| 1944_345.qff | | | | | | | | | | | | | | 1 | | | | | | | | i i i |
| | | | | | | 11 | | | <u>,</u> | | | 1 | 14 | 1 | | | 1 | | | | | 1. |
| 12. 19. 19. 19. 19. 19. 19. | 58_585.0 ⁴¹ 59_595.0 ⁴¹ 41_572.0 ⁴¹ 43_593.0 ⁴¹ 43_514.0 ⁴¹ | 57_554gtr 18_585gtr 11_552gtr 11_552 | orjasaji | nganak 11 8 11 Kanak 11 8 11 | ngana 11 0 11 1 1 Ngana 11 0 11 1 Ngana 11 0 11 1 Ngana 11 0 10 1 Ngana 11 0 10 1 Ngana 11 0 10 1 Ngana 11 0 10 1 | rganar 11 11 11 11 11 1 Konar 11 11 11 11 11 1 Konar 11 11 11 11 11 1 Konar 11 11 11 11 1 | rganar I.I. I. II. II. I. I. I. Konar I.I. II. II. II. I. I. Konar I.I. II. II. II. I. Konar I.I. II. II. I. | rganar II II II II I I II I Konar II II II II I I Norad II II II II I I Konar II II II II I Konar II II II I I Konar II II II II III | Implement Implement <t< td=""><td>rganar I.I. II. II. II. I. I.I. I.I. I. Konar I.I. II. II. II. I. I.I.I. I. Konar I.I. II. II. II. I. I.I.I. I. Konar I.I. II. II. I. I.I.I.</td><td>nganah 11 11 11 11 1 1 1 1 1 1 1 1 Kamah 11 11 11 11 1 1 1 1 1 1 1 Namah 11 11 11 1 1 1 1 1 1 1 Namah 11 11 11 1 1 1 1 1 1 Namah 11 11 11 1 1 1 1 1 1 Namah 11 11 11 1 1 1 1 1 1 Namah 11 11 11 1 1 1 1 1 Namah 11 11 11 1 1 1 1 1</td><td>rigendar I.I. II. II. II. I. I.I.I. I. I.I. nigendar I.I. II. II. II. I. I.I.I. I. I.I. nigendar I.I. II. II. II. I. I.I.I. I.I. nigendar I.I. II. II. II. I.I.I. I.I.I. nigendar I.I. II. II. II. I.I.I.I. I.I.I. nigendar I.I. II. II. II. I.I.I.I.I.I.I.I.I. nigendar I.I. II. II. II.I.I.I.I.I.I.I.I.I.I.I.</td><td>product 11 11 11 1 11 1 11 <t< td=""><td>product 11 11 11 1 11 <th< td=""><td>rgsmath 11 11 11 1 11 1 11 <!--</td--><td>rgsmap [1]<</td><td>rgsmap 11 10 11 1 11 <th1< th=""> <th1< th=""> 1<</th1<></th1<></td><td>product 11 11 11 1 11 1 11 1 11 1 11 1 11 1 11 1 11 1 11</td></td></th<><td>rgsmath 11 11 1 11 1</td><td>rgssade F1 F1</td><td>Special I<!--</td--><td>rised 1</td></td></td></t<></td></t<> | rganar I.I. II. II. II. I. I.I. I.I. I. Konar I.I. II. II. II. I. I.I.I. I. Konar I.I. II. II. II. I. I.I.I. I. Konar I.I. II. II. I. I.I.I. | nganah 11 11 11 11 1 1 1 1 1 1 1 1 Kamah 11 11 11 11 1 1 1 1 1 1 1 Namah 11 11 11 1 1 1 1 1 1 1 Namah 11 11 11 1 1 1 1 1 1 Namah 11 11 11 1 1 1 1 1 1 Namah 11 11 11 1 1 1 1 1 1 Namah 11 11 11 1 1 1 1 1 Namah 11 11 11 1 1 1 1 1 | rigendar I.I. II. II. II. I. I.I.I. I. I.I. nigendar I.I. II. II. II. I. I.I.I. I. I.I. nigendar I.I. II. II. II. I. I.I.I. I.I. nigendar I.I. II. II. II. I.I.I. I.I.I. nigendar I.I. II. II. II. I.I.I.I. I.I.I. nigendar I.I. II. II. II. I.I.I.I.I.I.I.I.I. nigendar I.I. II. II. II.I.I.I.I.I.I.I.I.I.I.I. | product 11 11 11 1 11 1 11 1 <t< td=""><td>product 11 11 11 1 11 <th< td=""><td>rgsmath 11 11 11 1 11 1 11 <!--</td--><td>rgsmap [1]<</td><td>rgsmap 11 10 11 1 11 <th1< th=""> <th1< th=""> 1<</th1<></th1<></td><td>product 11 11 11 1 11 1 11 1 11 1 11 1 11 1 11 1 11 1 11</td></td></th<><td>rgsmath 11 11 1 11 1</td><td>rgssade F1 F1</td><td>Special I<!--</td--><td>rised 1</td></td></td></t<> | product 11 11 11 1 11 <th< td=""><td>rgsmath 11 11 11 1 11 1 11 <!--</td--><td>rgsmap [1]<</td><td>rgsmap 11 10 11 1 11 <th1< th=""> <th1< th=""> 1<</th1<></th1<></td><td>product 11 11 11 1 11 1 11 1 11 1 11 1 11 1 11 1 11 1 11</td></td></th<> <td>rgsmath 11 11 1 11 1</td> <td>rgssade F1 F1</td> <td>Special I<!--</td--><td>rised 1</td></td> | rgsmath 11 11 11 1 11 1 11 1 </td <td>rgsmap [1]<</td> <td>rgsmap 11 10 11 1 11 <th1< th=""> <th1< th=""> 1<</th1<></th1<></td> <td>product 11 11 11 1 11 1 11 1 11 1 11 1 11 1 11 1 11 1 11</td> | rgsmap [1]< | rgsmap 11 10 11 1 11 1 <th1< th=""> <th1< th=""> 1<</th1<></th1<> | product 11 11 11 1 11 1 11 1 11 1 11 1 11 1 11 1 11 1 11 | rgsmath 11 11 1 11 1 | rgssade F1 F1 | Special I </td <td>rised 1</td> | rised 1 |

Conclusions

While low-pass WGS is an effective technical alternative to microarray-based genotyping, the economics of using sequencing in place of microarrays depends on the ability to reliably create balanced pools of libraries that efficiently multiplex as many genomic samples as possible on large high-output sequencing runs that achieve the lowest cost per unit of data.

Our results demonstrate the use of plexWell for routine low-pass WGS applications, where we show robust and scalable genotype imputation accuracy on large sample populations. The normalization capability of the plexWell workflow permits efficient generation of multiplexed libraries from hundreds to thousands of samples while avoiding sample dropouts due to input DNA variation.

References

Fragoso, Christopher, Maria Moreno, Zuoheng Wang, Christopher Heffelfinger, Lady J. Arbelaez, John A. Aguirre, Natalia Franco et al. "Genetic Architecture of a Rice Nested Association Mapping Population." G3: Genes, Genomes, Genetics (2017): g3-117.



