

Supplementary Material

Genomic prediction for targeted populations of environments in oat (*Avena sativa*)

Pablo Sandro^A, Madhav Bhatta^{A,B}, Alisha Bower^C, Sarah Carlson^C, Jean-Luc Jannink^D, David J. Waring^D, Clay Birkett^D, Kevin Smith^E, Jochum Wiersma^E, Melanie Caffè^F, Jonathan Kleinjan^F, Michael S. McMullen^G, Lydia English^C, and Lucia Gutierrez^{A,}*

^ADepartment of Plant and Agroecosystem Sciences, University of Wisconsin – Madison, 1575 Linden Drive, Madison, WI 53706, USA.

^BBayer Crop Science, Chesterfield, MO 63017, USA.

^CPractical Farmers of Iowa, 1615 Golden Aspen Drive, Suite 101, Ames, IA, USA.

^DSchool of Integrative Plant Science Plant Breeding and Genetics Section, Cornell University, 258 Emerson Hall, Ithaca, NY 14853, USA.

^EDepartment of Agronomy and Plant Genetics, University of Minnesota, 1991 Upper Buford Circle, 411 Borlaug Hall, St. Paul, MN 55108, USA.

^FDepartment of Agronomy, Horticulture, and Plant Sciences, South Dakota State University, Research Parkway, University Station, Brookings, SD 57006, USA.

^GPlant Sciences, North Dakota State University, Loftsgard 370D Dept 7670, PO Box 6050 Fargo, ND 58108-6050, USA.

*Correspondence to: Lucia Gutierrez Department of Plant and Agroecosystem Sciences, University of Wisconsin – Madison, 1575 Linden Drive, Madison, WI 53706, USA Email: gutierrezcha@wisc.edu

Table S1. Summary statistics and characterization of the 59 locations where oat performance was evaluated in 9 states of the Midwest of the United States from 1997 to 2021. Location, State, environmental code (Env), latitude (Lat), longitude (Lon), grain yield mean (g m^{-2}), the number of data-points in the dataset (N), mega-environment classification (ME, four MEs were obtained: me1 to me4), number of years (#year), first (first year), and last (last year) year in the dataset.

| State | Location | Env | Lat | Lon | grain yield (g m^{-2}) | N | ME | #years | first year | last year |
|-------|----------------|--------|-------|--------|-----------------------------------|------|-----|--------|------------|-----------|
| IA | Ames | IA_AME | 42.03 | -93.61 | 501 | 905 | ME2 | 12 | 1997 | 2008 |
| IA | Boone | IA_BON | 42.06 | 93.89 | 331 | 86 | ME3 | 3 | 2018 | 2020 |
| IA | Charles City | IA_CHA | 43.07 | -92.67 | 261 | 62 | ME3 | 4 | 2016 | 2019 |
| IA | Crawfordsville | IA_CRA | 41.21 | -91.53 | 437 | 235 | ME3 | 7 | 1997 | 2003 |
| IA | Kanawha | IA_KAN | 42.93 | -93.81 | 316 | 132 | ME3 | 6 | 2015 | 2020 |
| IA | Lewis | IA_LEW | 41.31 | -95.08 | 418 | 236 | ME3 | 7 | 1997 | 2003 |
| IA | Nashua | IA_NAS | 42.95 | -92.53 | 446 | 369 | ME3 | 13 | 1997 | 2020 |
| IL | Dekalb | IL_DKL | 41.93 | -88.75 | 391 | 242 | ME3 | 13 | 2002 | 2015 |
| IL | Monmouth | IL_MON | 40.92 | -90.64 | 452 | 434 | ME3 | 17 | 2004 | 2021 |
| IL | Urbana | IL_URB | 40.11 | -88.21 | 421 | 1698 | ME3 | 25 | 1997 | 2021 |
| IN | West Lafayette | IN_WES | 40.46 | -86.94 | 496 | 995 | ME1 | 17 | 1997 | 2013 |
| MI | Calumet | IA_CAL | 42.96 | -95.57 | 397 | 203 | ME2 | 6 | 1997 | 2003 |
| MI | East Lansing | MI_EAS | 42.75 | -84.47 | 360 | 796 | ME3 | 16 | 1997 | 2013 |
| MN | Crookston | MN_CRO | 47.76 | -96.58 | 579 | 1583 | ME2 | 8 | 2014 | 2021 |
| MN | Fergus Falls | MN_FER | 46.28 | -96.07 | 479 | 163 | ME1 | 5 | 2021 | 2016 |
| MN | Kimball | MN_KIM | 45.40 | -94.32 | 543 | 51 | ME3 | 2 | 2016 | 2019 |
| MN | Lamberton | MN_LAM | 44.23 | -95.28 | 438 | 1635 | ME3 | 6 | 2016 | 2021 |

| | | | | | | | | | | |
|----|--------------|---------|-------|-------------|-----|------|-----|----|------|------|
| MN | Lecenter | MN_LEC | 44.39 | -93.73 | 567 | 188 | ME4 | 6 | 2016 | 2021 |
| MN | Morris | MN_MOR | 45.58 | -95.91 | 448 | 1152 | ME2 | 19 | 1997 | 2020 |
| MN | Rochester | MN_ROCH | 44.01 | -92.47 | 530 | 90 | ME3 | 2 | 2020 | 2021 |
| MN | Roseau | MN_ROS | 48.82 | -95.75 | 617 | 244 | ME2 | 6 | 2016 | 2021 |
| MN | Rosemount | MN_RSM | 44.73 | -93.13 | 309 | 838 | ME2 | 15 | 1997 | 2011 |
| MN | Saint Paul | MN_SAI | 44.96 | -93.08 | 395 | 256 | ME3 | 3 | 2012 | 2020 |
| MN | Stephen | MN_STE | 48.48 | -96.83 | 562 | 188 | ME2 | 6 | 2016 | 2021 |
| MN | Waseca | MN_WAS | 44.07 | -93.51 | 355 | 632 | ME3 | 20 | 1999 | 2021 |
| ND | Carrington | ND_CAR | 47.45 | -99.09 | 426 | 708 | ME2 | 20 | 1998 | 2020 |
| ND | Dickinson | ND_DIC | 46.82 | - 102.66 | 432 | 282 | ME2 | 15 | 2005 | 2020 |
| ND | Edgeley | ND_EDG | 46.36 | -98.72 | 477 | 86 | ME2 | 5 | 2014 | 2018 |
| ND | Fargo | ND_FAR | 46.90 | -96.80 | 551 | 1066 | ME2 | 22 | 1997 | 2020 |
| ND | Hettinger | ND_HET | 46.04 | - 102.60 | 380 | 282 | ME3 | 15 | 2005 | 2020 |
| ND | Lamoure | ND_LAM | 46.36 | 98.29 | 425 | 28 | ME3 | 2 | 2012 | 2013 |
| ND | Langdon | ND_LNG | 48.81 | -98.31 | 575 | 274 | ME2 | 15 | 2005 | 2020 |
| ND | Minot | ND_MIN | 48.23 | - 101.27 | 454 | 958 | ME1 | 23 | 1997 | 2020 |
| ND | Williston | ND_WIL | 48.17 | - 103.61 | 285 | 223 | ME2 | 12 | 2005 | 2020 |
| NY | Batavia | NY_BAT | 43.00 | -78.19 | 253 | 80 | ME3 | 4 | 2015 | 2018 |
| NY | Henrietta | NY_HEN | 43.06 | 77.61 | 199 | 53 | ME3 | 2 | 2009 | 2011 |
| NY | Ithaca | NY_ITH | 42.45 | -76.50 | 322 | 1257 | ME2 | 23 | 1997 | 2020 |
| NY | Sprout Brook | NY_SPR | 42.85 | 74.69 | 255 | 38 | ME3 | 2 | 2013 | 2014 |
| NY | Wayland | NY_WAY | 42.56 | -77.60 | 248 | 151 | ME1 | 7 | 2011 | 2018 |
| SD | Beresford | SD_BER | 43.07 | -96.79 | 379 | 910 | ME3 | 21 | 1999 | 2020 |

| | | | | | | | | | | |
|----|----------------|----------|-------|-------------|-----|------|-----|----|------|------|
| SD | Bison | SD_BIS | 45.52 | - 102.46 | 196 | 28 | ME2 | 2 | 2014 | 2015 |
| SD | Brookings | SD_BRO | 44.31 | -96.79 | 415 | 1667 | ME3 | 23 | 1997 | 2020 |
| SD | Draper | SD_DRP | 43.86 | - 100.53 | 308 | 29 | ME2 | 2 | 2016 | 2017 |
| SD | Miller | SD_MIL | 44.52 | -98.99 | 375 | 86 | ME2 | 6 | 2014 | 2020 |
| SD | Okaton | SD_OKAT | 43.96 | - 100.92 | 317 | 42 | ME2 | 3 | 2014 | 2020 |
| SD | Selby | SD_SELBY | 45.51 | - 100.03 | 503 | 71 | ME1 | 6 | 2014 | 2021 |
| SD | South Shore | SD_SOU | 45.13 | -97.03 | 483 | 1037 | ME3 | 21 | 1997 | 2020 |
| SD | Wall | SD_WAL | 43.99 | - 102.24 | 299 | 72 | ME2 | 5 | 2014 | 2020 |
| SD | Winner | SD_WIN | 43.38 | -99.86 | 297 | 85 | ME2 | 6 | 2015 | 2020 |
| WI | Antigo | WI_ANT | 45.12 | -89.13 | 267 | 84 | ME2 | 4 | 2018 | 2021 |
| WI | Arlington | WI_ARL | 43.32 | -89.37 | 456 | 1340 | ME3 | 22 | 1998 | 2021 |
| WI | Ashfield | WI_ASH | 44.50 | -90.00 | 366 | 106 | ME2 | 4 | 1998 | 2001 |
| WI | Buffalo County | WI_BUF | 44.47 | -91.68 | 346 | 78 | ME3 | 4 | 2018 | 2020 |
| WI | Chilton | WI_CHI | 44.03 | -88.18 | 322 | 503 | ME3 | 18 | 1997 | 2019 |
| WI | Lancaster | WI_LAN | 42.85 | -90.70 | 362 | 541 | ME3 | 19 | 1997 | 2021 |
| WI | Madison | WI_MAD | 43.07 | -89.42 | 403 | 3242 | ME3 | 24 | 1997 | 2021 |
| WI | Marshfield | WI_MAR | 44.47 | -90.02 | 358 | 536 | ME2 | 18 | 1998 | 2017 |
| WI | Spooner | WI_SPO | 45.85 | -91.94 | 268 | 607 | ME3 | 21 | 1997 | 2020 |
| WI | Sturgeon Bay | WI_STU | 44.84 | -87.38 | 383 | 610 | ME3 | 21 | 1997 | 2018 |

Table S2. General description of genomic prediction models and strategies used in the characterization of a large oat population evaluated in 59 locations over a period of 25 years in the Midwest United States. Models are described as modification from model [1] $y_{ij}=\mu+g_i+e_j+ge_{ij}+\epsilon_{ij}$, where genotype by environment interaction was modeled either as genotype (G) by location (LOC), year (YEAR), or mega-environment (ME). The variance-covariance matrix for the genotype by environment interaction was modeled in general as the Kronecker product of the realized additive relationship matrix estimated with marker data ($\mathbf{K}_{(g \times g)}$) and the genetic variance-covariance among environments ($\Sigma_{\mathbf{E}(exe)}$), with $\Sigma_{\mathbf{E}}$ either an independent variance-covariance structure, a factor-analytic (FA(1)), a diagonal matrix (\mathbf{D}), or a known unstructured (UN*) variance-covariance. The residuals were modeled with a variance-covariance matrix, $\mathbf{R}_{([gxe] \times [gxe])}$, as a diagonal matrix with the reciprocals of the standard error of the BLUEs estimates for each genotype in each environment to account for the heterogeneity in the precision of means estimates in the two-step model following Piepho et al. (2012).

| Strategy | y | Fixed | GEI | Other Components | Res | Var(ge_{ij}) | $\Sigma_{\mathbf{G}}$ | $\Sigma_{\mathbf{E}}$ | Var(ϵ_{ij}) |
|---|---------------------|---------------------|---|---|-----------------|---|---|--------------------------------------|-------------------------------|
| <i>GEI Characterization</i> | | | | | | | | | |
| VC | gxlxa | mu | $ge_{ij}=\mathbf{G} \cdot \mathbf{LOC} \cdot \mathbf{YEAR}$ | | ϵ_{ij} | \mathbf{I} | | | $\mathbf{R}=\mathbf{I}(1/SE)$ |
| GGE | gxl | $e_j=\mathbf{LOC}$ | $ge_{ij}=\underline{\mathbf{G}} \cdot \mathbf{LOC}$ | | ϵ_{ij} | $\Sigma_{\mathbf{G}} \otimes \Sigma_{\mathbf{E}}$ | $\Sigma_{\mathbf{G}}=\sigma_g^2 \mathbf{K}$ | $\Sigma_{\mathbf{E}}=\mathbf{FA}(1)$ | $\mathbf{R}=\mathbf{I}(1/SE)$ |
| <i>Un-phenotyped individuals</i> | | | | | | | | | |
| CV1 by \mathbf{ME}_m | gxl _m | $e_j=\mathbf{LOC}$ | $ge_{ij}=\mathbf{G} \cdot \mathbf{LOC}$ | <u>\mathbf{G}</u> | ϵ_{ij} | $\Sigma_{\mathbf{G}} \otimes \Sigma_{\mathbf{E}}$ | $\Sigma_{\mathbf{G}}=\sigma_g^2 \mathbf{K}$ | $\Sigma_{\mathbf{E}}=\mathbf{D}$ | $\mathbf{R}=\mathbf{I}(1/SE)$ |
| CV1 Overall | gxl | $e_j=\mathbf{ME}$ | $ge_{ij}=\mathbf{G} \cdot \mathbf{ME}$ | <u>\mathbf{G}</u> | ϵ_{ij} | $\Sigma_{\mathbf{G}} \otimes \Sigma_{\mathbf{E}}$ | $\Sigma_{\mathbf{G}}=\sigma_g^2 \mathbf{K}$ | $\Sigma_{\mathbf{E}}=\mathbf{D}$ | $\mathbf{R}=\mathbf{I}(1/SE)$ |
| <i>Leave one location out</i> | | | | | | | | | |
| CV0 _{LOC M} by \mathbf{ME}_m | gxl _m | $e_j=\mathbf{LOC}$ | $ge_{ij}=\underline{\mathbf{G}} \cdot \mathbf{LOC}$ | | ϵ_{ij} | $\Sigma_{\mathbf{G}} \otimes \Sigma_{\mathbf{E}}$ | $\Sigma_{\mathbf{G}}=\sigma_g^2 \mathbf{K}$ | $\Sigma_{\mathbf{E}}=\mathbf{FA}(1)$ | $\mathbf{R}=\mathbf{I}(1/SE)$ |
| CV0 _{LOC M} Overall | gxl | $e_j=\mathbf{ME}$ | $ge_{ij}=\mathbf{G} \cdot \mathbf{ME}$ | $\mathbf{ME} \cdot \mathbf{LOC} + \underline{\mathbf{G}}$ | ϵ_{ij} | $\Sigma_{\mathbf{G}} \otimes \Sigma_{\mathbf{E}}$ | $\Sigma_{\mathbf{G}}=\sigma_g^2 \mathbf{K}$ | $\Sigma_{\mathbf{E}}=\mathbf{FA}(1)$ | $\mathbf{R}=\mathbf{I}(1/SE)$ |
| CV0 _{LOC GEI} by \mathbf{ME}_m | gxl _m | $e_j=\mathbf{LOC}$ | $ge_{ij}=\underline{\mathbf{G}} \cdot \mathbf{LOC}$ | | ϵ_{ij} | $\Sigma_{\mathbf{G}} \otimes \Sigma_{\mathbf{E}}$ | $\Sigma_{\mathbf{G}}=\sigma_g^2 \mathbf{K}$ | $\Sigma_{\mathbf{E}}=\mathbf{UN}^*$ | $\mathbf{R}=\mathbf{I}(1/SE)$ |
| <i>Leave one year out</i> | | | | | | | | | |
| CV0 _{Y ME} by \mathbf{ME}_m | gxl _m xa | $e_j=\mathbf{ME}$ | $ge_{ij}=\underline{\mathbf{G}} \cdot \mathbf{ME}$ | $\mathbf{YEAR} \mathbf{LOC} \mathbf{ME}$ | ϵ_{ij} | $\Sigma_{\mathbf{G}} \otimes \Sigma_{\mathbf{E}}$ | $\Sigma_{\mathbf{G}}=\sigma_g^2 \mathbf{K}$ | $\Sigma_{\mathbf{E}}=\mathbf{FA}(1)$ | $\mathbf{R}=\mathbf{I}(1/SE)$ |
| CV0 _{Y ME} Overall | gxlxa | $e_j=\mathbf{ME}$ | $ge_{ij}=\mathbf{G} \cdot \mathbf{ME}$ | $\mathbf{YEAR} \mathbf{LOC} \mathbf{ME} + \underline{\mathbf{G}}$ | ϵ_{ij} | $\Sigma_{\mathbf{G}} \otimes \Sigma_{\mathbf{E}}$ | $\Sigma_{\mathbf{G}}=\sigma_g^2 \mathbf{K}$ | $\Sigma_{\mathbf{E}}=\mathbf{FA}(1)$ | $\mathbf{R}=\mathbf{I}(1/SE)$ |
| CV0 _{Y LOC} by \mathbf{LOC}_m | gxa _m | $e_j=\mathbf{YEAR}$ | $ge_{ij}=\underline{\mathbf{G}} \cdot \mathbf{YEAR}$ | | ϵ_{ij} | $\Sigma_{\mathbf{G}} \otimes \Sigma_{\mathbf{E}}$ | $\Sigma_{\mathbf{G}}=\sigma_g^2 \mathbf{K}$ | $\Sigma_{\mathbf{E}}=\mathbf{FA}(1)$ | $\mathbf{R}=\mathbf{I}(1/SE)$ |
| <i>Predicting years</i> | | | | | | | | | |
| YEAR by \mathbf{ME}_m | gxl _m xa | $e_j=\mathbf{LOC}$ | $ge_{ij}=\underline{\mathbf{G}} \cdot \mathbf{LOC}$ | \mathbf{YEAR} | ϵ_{ij} | $\Sigma_{\mathbf{G}} \otimes \Sigma_{\mathbf{E}}$ | $\Sigma_{\mathbf{G}}=\sigma_g^2 \mathbf{K}$ | $\Sigma_{\mathbf{E}}=\mathbf{FA}(1)$ | $\mathbf{R}=\mathbf{I}(1/SE)$ |

| | | | | | | | | | |
|------------------------------|---------------------|--------------------|---------------------------------------|-------------------------------|-----------------|------------------------------------|--|----------------------------|------------------|
| YEAR Overall | gxlxa | e _j =ME | ge _{ij} = <u>G·ME</u> | ME·LOC + YEAR | ε _{ij} | Σ_G⊗Σ_E | Σ_G=σ_g²K | Σ_E=I | R=I(1/SE) |
| <i>Variety decision tool</i> | | | | | | | | | |
| VDT by ME _m | gxl _m xa | e _j =ME | ge _{ij} = <u>G·ME</u> | YEAR LOC ME | ε _{ij} | Σ_G⊗Σ_E | Σ_G=σ_g²K | Σ_E=FA(1) | R=I(1/SE) |
| VDT Overall | gxlxa | e _j =ME | ge _{ij} = G·ME | YEAR LOC ME + <u>G</u> | ε _{ij} | Σ_G⊗Σ_E | Σ_G=σ_g²K | Σ_E=FA(1) | R=I(1/SE) |

Table S3. Predictive ability within and across mega-environments (ME) for different cross validation strategies. The $CV0_{LOC|GEI}$ used a leave-one-location-out approach with a GEI GBLUP modeling approach ($GBLUP_{GxE}$) within each ME. The $CV0_{LOC|M}$ uses a leave-one-location-out approach with an average GBLUP modeling approach ($GBLUP_M$) within each ME. The $CV0_{Y|LOC}$ used a leave-one-year-out approach with a $GBLUP_M$ modeling approach within each location. The $CV0_{Y|ME}$ used a leave-one-year-out approach with a $GBLUP_M$ modeling approach within each ME. Finally, the CV1 used a 10-fold cross-validation strategy for predicting the performance of genotypes that have not been evaluated in any environment within a ME (ME1-ME3) or overall.

| | ME1 | ME2 | ME3 | Overall |
|-----------------|------------------|------------------|------------------|------------------|
| CV1 | 0.64 (0.02-0.67) | 0.45 (0.01-0.47) | 0.65 (0.01-0.67) | 0.57 (0.41-0.69) |
| $CV0_{LOC M}$ | 0.27 (0.15-0.51) | 0.55 (0.27-0.89) | 0.55 (0.15-0.82) | -- |
| $CV0_{LOC GEI}$ | 0.36 (0.19-0.62) | 0.47 (0.24-0.86) | 0.55 (0.14-0.87) | -- |
| $CV0_{Y LOC}$ | 0.52 (0.10-0.83) | 0.51 (0.05-0.95) | 0.53 (0.20-0.96) | -- |
| $CV0_{Y ME}$ | 0.30 (0.55-0.10) | 0.24 (0.62-0.10) | 0.25 (0.50-0.02) | -- |
| #genotypes | 304 | 393 | 401 | 401 |
| #locations | 5 | 23 | 30 | 59 |