nature research

Corresponding author(s): Masakazu Iwai, Krishna K. Niyogi

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☒	☐	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☒	☐	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☒	☐	A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated
☒	☐	Clearly defined error bars *State explicitly what error bars represent (e.g. SD, SE, CI)*

Our web collection on statistics for biologists may be useful.

Software and code

Policy information about availability of computer code

Data collection	We used Gautomatch (version 0.53, http://www.mrc-lmb.cam.ac.uk/kzhang/) to automatically select protein particles for single-particle analysis. For MS data acquisition, MassLynx software (version 4.1) was used.
Data analysis	We used RELION (version 1.4) to analyze the data for single-particle analysis. For MS data analysis, Protein Lynx Global Server software (version 3.0.2, Waters) was used.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The data are available from the corresponding authors (M.I. and K.K.N.) upon request.

Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/authors/policies/ReportingSummary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Sample size was determined by the requirement for the data analysis to produce reliable reconstructed structures. We collected the protein particles ranging from 19,000 to 56,000, which are sufficient to obtain reliable structures from single-particle analysis. All the data set sizes are summarized in Supplementary Table 2.
Data exclusions	To improve the quality of the resolution and remove the bad or non-related structures, we excluded some of the data. As described in Supplementary Information, we performed the first classification using all the data and removed bad classes. Using good classes, we refined the data to improve the quality. This method has been described previously and pre-established for single-particle analysis. The details for classification and refinement were summarized in Supplementary Information.
Replication	All the data we presented in this study were successfully reproduced.
Randomization	Randomization is not relevant to our study because we tried to identify the specific structural organizaton of the protein supercomplexes through the purification using specific procedures as described in Methods section.
Blinding	Blinding is not relevant to our study because we tried to identify the specific structural organizaton of the protein supercomplexes through the purification using specific procedures as described in Methods section.

Reporting for specific materials, systems and methods

Materials & experimental systems

n/a	Involved in the study
☐	☒ Unique biological materials
☐	☒ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☒	☐ Human research participants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Unique biological materials

Policy information about availability of materials

Obtaining unique materials	All unique materials used are readily available from the authors.

Antibodies

Antibodies used	Antibodies specific for PsaD (AS09 461, 1/5000 dilution), Lhca1 (AS01 005, 1/5000 dilution), Lhcb2 as Lhcbm (AS01 003, 1/5000 dilution), and Lhcb4 (AS04 045, 1/2000 dilution) were purchased from Agrisera (www.agrisera.com). The antibody specific for Lhcb9.1 was raised against the synthetic peptide with the sequences CQAVSKGEGTADAPT by Eurofins Genomics (www.eurofinsgenomics.com). Lhcb9.1 antibody was used in 1/5000 dilution.
Validation	The specificity of each antibody to Physcomitrella patens proteins has been confirmed by the previous study and described in the Supplementary Information (Iwai et al., Nature Plants, 2015:1, 14008). All the antibodies were tested on 2DE and the corresponding protein spots were identified by mass spectrometry.