PHYLOGENETIC ANALYSIS OF PHOTOSYSTEMS I AND II

Steven C Halls, Stephen R Downie, and Govindjee

University of Illinois, Urbana, IL 61801 U.S.A. E-mail: gov@uiuc.edu

Photosystem I (PS I) and Photosystem II (PS II) reaction centers include heterodimers of the gene products of *psa*A (PSI-A) and *psa*B (PSI-B) and of *psb*A (D1) and *psb*D (D2), respectively. Structural and sequence similarity exists among the heterodimer pair PSI-A and PSI-B, and D1 and D2 (Svensson et al., 1990; Golbeck and Bryant, 1991; Nitschke and Rutherford, 1991; Blankenship, 1992; Ruffle et al., 1992). Many sequences are now available for reaction center proteins of PS I and PS II. Golbeck and Bryant (1991) presented alignments of PSI-A and PSI-B for 7 taxa. Svensson et al. (1992) compiled 38 sequences of D1 (27 taxa) and 15 sequences of D2 (12 taxa) and showed that the sequences were strongly conserved. Ruffle et al. (1992) have shown relationships among the D1 and D2 proteins. In this paper, we present the first phylogenetic analysis of PSI-A and PSI-B, and extend the existing reaction center alignments of D1 and D2 to include several additional taxa.

Selected published protein sequences were retrieved from SWISS-PROT, PIR, and GenPept data bases using BLAST (Altschul et al., 1990). Additional sequences were obtained for *Porphyra purpurea* (M. Reith and J. Munholland, unpublished), *Cyanophora paradoxa* (V. L. Stirewelt and D. A. Bryant, unpublished), and *Zea mays* (Sopory et al., 1993, from Larrinua and McLaughlin, 1987). In total, 39 D1 and 11 D2 sequences of different genera, and 14 PSI-A and 14 PSI-B sequences, representing 13 genera each, were aligned. Alignments were done as described by Svensson et al. (1992), and Golbeck and Bryant (1991). The resulting data matrices were analyzed using PAUP (Phylogenetic Analysis Using Parsimony) version 3.1.1 (Swofford, 1993). All heuristics searches were replicated 100 times with random addition sequence and tree bisection-reconnection branch swapping. The options mulpars, steepest descent, collapse, and acctran optimization were selected. Gaps were treated as missing data. The amount of phylogenetic information in the parsimony analyses was estimated using the consistency, CI, (excluding uninformative characters) and retention, RI, indices. The trees were rooted by positioning the root along the branch connecting *Synechococcus* to the rest of the network.

Multiple alignments of D1, D2, PSI-A, and PSI-B sequences resulted in matrices of 360, 353, 804, and 743 positions, respectively; consensus sequences for each group of reaction center proteins are presented below (Complete data matrices are available upon request.).

```
D1 (Photosystem II)

PROKARYOTES   MTTTLERRES ASLWERFCSW ITSTDERLYV GWFGVLMIPT LLAATICFII AFIAAPPVDI DGIREPVAGS LLYGNNIISG AVVPSSNAIG LHFYPIWEAA
ALGAE           AI           A   E                          T  SV                      S           T    I T
MONOCOTS        AI       T   G   N                          T  SV                      S               II T A
DICOTS          AI       E   G   N                          T  SV                      S                I T A

PROKARYOTES   SLDEWLYNGG PYQLVVFHFL IGICCYMGRE WELSYRLGMR PWICVAYSAP VAAATAVFLI YPIGQGSFSD GMPLGISGTF NFMLVFQAEH NILMHPFHML
ALGAE                        I    F   V                      A  F        S    V                          I
MONOCOTS       V           E I       L VA                    A                                           I
DICOTS         V           E I       L VA                    A           F                               I

PROKARYOTES   GVAGVFGGSL FSAMHGSLVT SSLVRETTEN ESQNYGYKFG QEEETYNIVA AHGYFGRLIF QYASFNNSRS LHFFLAAWPV VGIWFTALGV STMAFNLNGF
ALGAE                                  I         A N
MONOCOTS                               I         A E  R                                                        I
DICOTS                                 I         A E  R                                                        I

PROKARYOTES   NFNQSVVDSQ GRVINTWADI INRANLGMEV MHERNAHNFP LDLASVESAP VALTAPSING
ALGAE                                                                --- ----
MONOCOTS                                                          A  --- ----
DICOTS                                                            A  --- ----
```


P. Mathis (ed.), Photosynthesis: from Light to Biosphere, Vol. I, 955–958.

D2 (Photosystem II)

```
PROKARYOTES     MTIAVGRAAA VERGWFDVLD DWLKRDRFVF IGWSGLLLFP CAFMALGGWL TGTTFVTSWY THGLASSYLE GANFLTVAVT SPADAFGHSL LFLWGPEAQN
ALGAE               I TTYK EK     DA                         SYF                       T      C    A  S T  NSMA      V
MONOCOTS            L  VPK E NDL  IM     R                    YF                              C    A  S T  NSLA     L
DICOTS              L  FTK D NDL  IM     R                    YF V                            C    A    T  NSLA     L

PROKARYOTES     NLTRWFQLGG LWFFVALHGA FGLIGFMLRQ FEIARLVGIR PYNAIAFSGP IAVFVSVFLM YPLGQSSWFF APSFGVAGIF RFILFLQGFH NWTLNPFHMM
ALGAE           DF   C I     A                          N           A           I                   A
MONOCOTS        DF   C       T         A           L    Q        S              I                   A
DICOTS          DF   C       T                     L    Q                       I                   A                    M

PROKARYOTES     GVAGILGGAL LCAIHGATVE NTLFEDGEDS NTFRAFEPTQ AEETYSMVTA NRFWSQIFGI AFSNKRWLHF FMLFVPVTGL WMSSVGIVGL ALNLRAYDFV
ALGAE                                        DGA       N    S                   V                          AI
MONOCOTS                                     DGA       N                        V                          AI V
DICOTS                                       DGA       N                        V                          AL V

PROKARYOTES     SQELRAAEDP EFETFYTKNI LLNEGIRAWM APQDQPHENF IFPEEVLPRG NAL
ALGAE              I                              A         V
MONOCOTS           I                              A       L
DICOTS             I                              A       L
```

PSI-A (Photosystem I)

```
PROKARYOTES     MTISPPEREA -KARVVVDND PVPTSFEKWA KPGHFDRTLA RGPKTTTWIW NLHALAHDFD THTSDLEDIS RKIFSAHFGH LAVVFIWLSG MYPHGAKFSN
ALGAE               T      K VKIA  K    E             S   S K  N           D      S            V      Q  GII
MONOCOTS         M RS  P-- -EVKI   R    K    E   R    S  L  K  D           D      S      E     V      Q  SII
DICOTS           I RS  P-- -EVKIL  R    K    E   R    S  I  K  D           D      S      E     V      Q  SII L

PROKARYOTES     YEAWLADPTG IKPSAQVVWP IVGQGILNGD VGGGFHGIQI TSGLFQLWRA SGITNSFQLY CTAIGGLVMA GLMLFAGWFH YHKRAPKLEW FQNVESMLNH
ALGAE                S                    E           Q        L           SEL    T          A  F          A
MONOCOTS             S      G             E           R     S  F  I        SEL         A IF  S             A    A  SHD
DICOTS               S      G             E           R        F  I        SEL         A  F  A             A    A    D

PROKARYOTES     HLAGLLGLGS LAWAGHQIHV SLPINKLLDA GVAAKDIPLP HEFILNPSLY PELDPGFASG LTPFFTLNWG AYSDFLTFKG GLNPVTGGLW LSDTAHHHLA
ALGAE             G                            S   DPQE            RD M AQ Y S SK   A       S E
MONOCOTS                    S              QF      DP E            RD L AQ Y S  ER A          K AE  S R      I         I
DICOTS                      S              QF N    DP-E        -   RD L AQ Y S  E  A          K A     R              T

PROKARYOTES     IAVLFIIAGH MYRTNWGIGH SLKEILEAHK GPFTGEGHKG LYEVLTTSWH AQLAINLAML GSLTIIVAQH MYAMPPYPYL ATDYPTQLSL FTHHMWIGGF
ALGAE             V  LV                M       R               I                F    S    H    S
MONOCOTS          I  L                G  D            QR       I          SL        T  V  H
DICOTS            I  L                G  D            Q        I           L           V  H

PROKARYOTES     LIVGGAAHGA IFMVRDYDPA MNQNNVLDRV LRHRDAIISH LNWVCIFLGF HSFGLYIHND TMRALGRPQD MFSDTAIQLQ PVFAQWVQNL HTLAPGGTAP
ALGAE           C   AG  A             N Y  L     I                                  S                         I  T  F    L
MONOCOTS            A   A           T TRY DL                                        S                               AG  RL
DICOTS              A   A           T TRY DL                    A                   S                         I  T  A   SA

PROKARYOTES     NALA-TASYA FGGDVVAVGG KVAMMPIVLG TADFMVHHIH AFTIHVTVLI LLKGVLFARS SRLIPDKANL GFRFPCDGPG RGGTCQVSGW DHVFLGLFWM
ALGAE               -    W G- ELG H          S    S
MONOCOTS        G TTS SLTW G  EL          LL  P       L
DICOTS          G TAS SLTW G S L          LL  P       L

PROKARYOTES     YNSLSIVIFH FSWKMQSDVW GTVAPDGTVS HITGGNFAQS AITINGWLRD FLWAQASQVI NSYGSALSAY GLMFLGAHFV WAFSLMFLFS GRGYWQELIE
ALGAE                V                   TAS G                              S     Q            I
MONOCOTS           I V                  ISDQ V T              S                   Q    S       F
DICOTS             I V                 SISDQ V T              S                   Q    S       F

PROKARYOTES     SIVWAHNKLK VAPAIQPRAL SIIQGRAVGV AHYLLGGIAT TWAFFLARIL SVG
ALGAE                                   T                     S      I A
MONOCOTS                       T                 T                   I A
DICOTS                         T        V        T                   I A
```

PSI-B (Photosystem I)

```
PROKARYOTES     MATKFPKFSQ DLAQDPTTRR IWYGIATAHD FETHDGMTEE NLYQKIFASH FGHLAIIFLW TSGNLFHVAW QGNFEQWVKD PVNTRPIAHA IWDPHFGKAA
ALGAE                      G                       S                     Q                          IT   LHV               P
MONOCOTS         ELR  R    G            F          S  DI    R   N        Q                        S IQ                     P
DICOTS            LR  R    G            F          S  DI    R   N        Q                           Q                     P

PROKARYOTES     VEAFTQAGAS NPVNIAYSGV YHWWYTIGMR TNGELYQGAI FLLILASLAL FAGWLHLQPK FRPSLSWFKN AESRLNHHLA GLFGVSSLAW AGHLVHVAIP
ALGAE                RG    G           Q           QD  T SV     VSA  F              Q                  S            T
MONOCOTS             RG  A G           Q      L    ED  T  L     F ST  S  IG         WK                 S            T
DICOTS               RG  L             Q      L    ED  T  L        AIS  I         WK   V              S            T

PROKARYOTES     ESRGQHVGWD NFLSTMPHPA GLAPFFTGNW GVYAQNPDTA SHVFGTSQGA GTAILTFLGG FHPQTESLWL TDIAHHHLAI AVLFIVAGHM YRTNFGIGHS
ALGAE               R          VL   Q       S    AA                E S  Q                      M                             R
MONOCOTS        A   EY R N     DVL   Q   G LL   Q  NL       SS N L   T                                      FI LI
DICOTS          A   EY R N     DVL       G L    Q  NL       SS   L   A                                      FI L

PROKARYOTES     IKEILNAHTK VGGPFGAGHQ GLYDTYNNSL HFQLGLALAS LGVITSLVAQ HMYSLPPYAF IAQDHTTQAA LYTHHQYIAG FLMVGAFAHG AIFLVRDYDP
ALGAE           L A  D  KP PS NL    K      I                                        I F                  I C          FI
MONOCOTS          DL E  TP P  RL R  K      I                                 S        F                  I T          FI
DICOTS          M DL D  IP P  RL R  K      I                                 A        F                                I

PROKARYOTES     AQNKGNVLAR VLDHKEAIIS HLSWVSLFLG FHTLGLYVHN DVVVAFGTPE KQILIEPVFA QFIQAAHGKL LYGFDVLLSN PDSIASTAWP NYGNVWLPGW
ALGAE           E          I               T                  MQ                            A          S STSV    GQ S---I
MONOCOTS        E  ED      M              A                   ML               I       S    T T    I   S TSGP FN GR T---L
DICOTS          E  ED      M E            A                   ML               I       S    T S        S TSGP FN GR S---I

PROKARYOTES     LDAINSGTNS LFLTIGPGDF LVHHAIALGL HTTTLILVKG ALDARGSKLM PDKKDFGYAF PCDGPGRGGT CDISAWDAFY LAMFWHLNTI GWVTFYWHWK
ALGAE                NNQ
MONOCOTS         N V ENS
DICOTS           N V ENS

PROKARYOTES     HLGVWEGNVA QFNESSTYLM GWFRDYLWLN SSQLINGYNP FGTNNLSVWA WMFLFGHLVW ATGFMFLISW RGYWQELIET LVWAHERTPL ANLVRWKDKP
ALGAE             TL                                                      M S      F                                      K
MONOCOTS         ITL QG   S                                               M S                                        A            I
DICOTS            TL QG   S                                               M S                                        A            I

PROKARYOTES     VALSIVQARL VGLAHFSVGY ILTYAAFLIA STAGKFG
ALGAE                                  F           S
MONOCOTS                               F           S
DICOTS                                 F           S
```

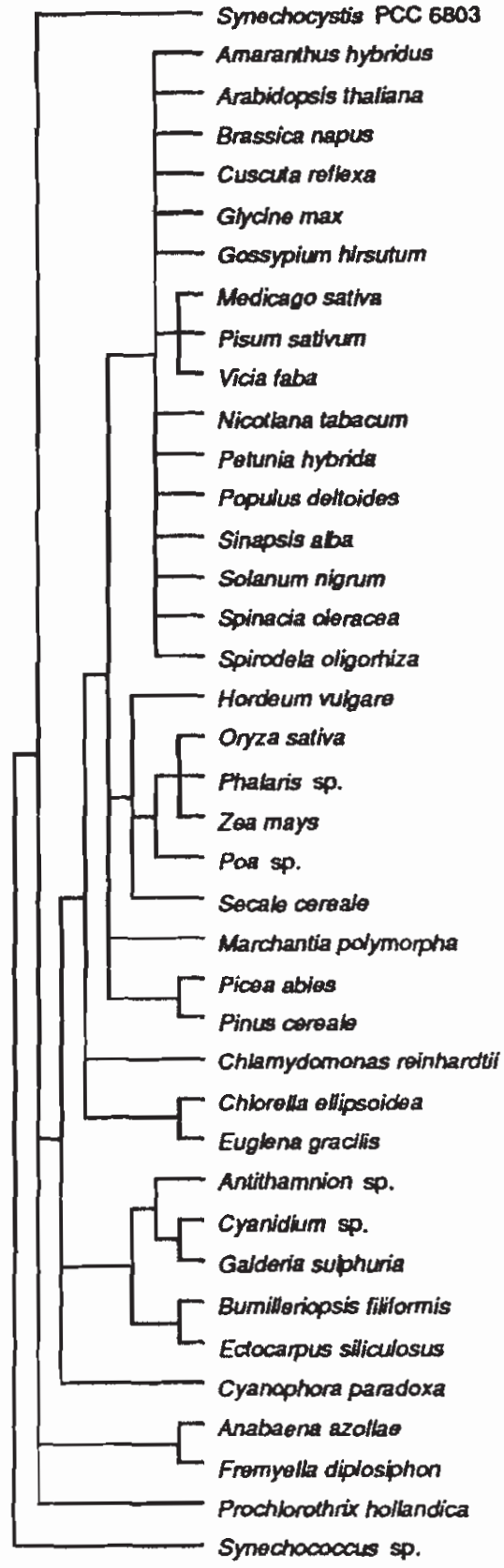

Fig. 1. Strict consensus of 1,660 minimal length 400-step trees derived from parsimony analysis of selected *psbA* (D1) sequences. CI=0.62, RI=0.69.

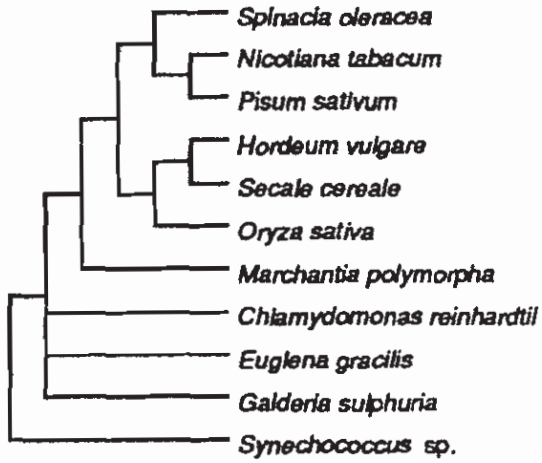

Fig. 2. Strict consensus of 4 minimal length 144-step trees derived from parsimony analysis of selected *psbD* (D2) sequences. CI=0.73, RI=0.62.

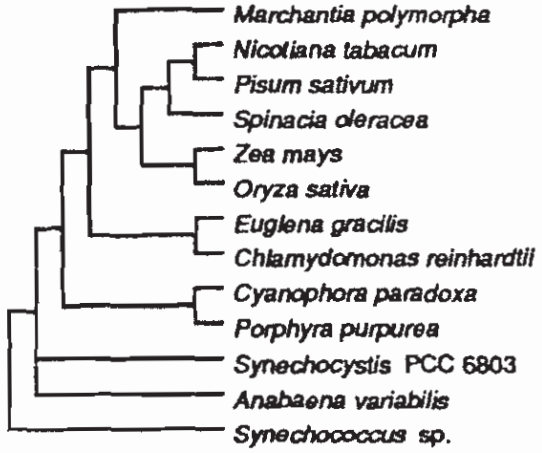

Fig. 3. Strict consensus of 2 minimal length 659-step trees derived from parsimony analysis of selected PSI-A sequences. CI=0.69, RI=0.69.

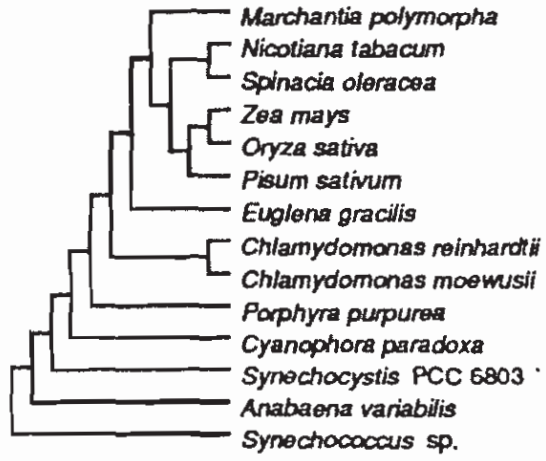

Fig. 4. Single maximally parsimonious 696-step tree derived from the analysis of selected PSI-B sequences. CI=0.72, RI=0.74.

The multiple alignment of each of the reaction center subunits was facilitated by the introduction of none (i.e., D2) or very few gaps (D1, one gap; PSI-B, three gaps; PSI-A, 16 gaps). Analyses of the D1, D2, and PSI-A data matrices resulted in 1,660, 4, and 2 maximally parsimonious topologies, whose strict consensus trees, with accompanying tree lengths, consistency and retention indices, are shown in Figs. 1, 2, and 3, respectively. A single maximally parsimonious tree obtained from the analysis of PSI-B sequences, is presented in Fig. 4. PS I and II reaction center proteins are highly conserved among the taxa examined. The inclusion of the *Pisum sativum* A2 PSI-A sequence (Lehmbeck et al., 1986) resulted in a single maximally-parsimonious tree of 1033 steps, 374 steps greater than the trees constructed without this sequence (CI excluding uninformative characters = 0.70; RI = 0.64). In this single PSI-A tree (not shown), this *Pisum* sequence arises as sister taxon to *Euglena* and away from the other *Pisum* and dicot sequences. In most cases, the evolutionary relationships inferred on the basis of PS I and II reaction center protein sequences and maximum parsimony are in agreement with traditional concepts of relationship. Generally the dicots and monocots each comprise a clade and are sistergroups.

Remarks: We have presented here the current consensus sequence of core PS II reaction center proteins (D1 and D2), and of core PS I reaction center proteins (PSI-A and PSI-B) based on published as well as unpublished sequences. Further, we have presented here the first phylogenetic analysis of PSI-A and PSI-B proteins and an extended analysis of D1 and D2 proteins. The phylogenetic trees, thus obtained, provide a hypothesis of the evolutionary relationship of the reaction centers of prokaryotes, cyanelle, rhodophytes, chlorophytes and other eukaryotes (gymnosperms, monocots and dicots).

Acknowledgments

Steven Halls acknowledges an IPR predoctoral training grant for financial support. We thank Drs. D. A. Bryant, V. L. Stirewelt, M. Reith and J. Munholland for providing us with unpublished sequences.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. and Lipman, D. J. (1990). *J. Mol. Biol.* **215**, 403-410.

Blankenship R. E. (1992). *Photosynth. Res.* **33**, 91-111.

Golbeck, J. H. and Bryant, D. A. (1991). *Current Topics in Bioenerg.* **16**, 83-176.

Larrinua, I. M. and McLaughlin, W. E. (1987). In: "*Progress in Photosynthesis Research*", edited by J Biggins, Vol 4, Nijhoff, The Hague, pp. 649-652.

Lehmbeck, J., Rasmusen, O. K., Bookjans, G. B., Jepsen, B. R., Stummann, B. M. and Henningsen, K. W. (1986). *Plant Mol. Biol.* **7**, 3-10.

Nitschke W. and Rutherford A. W. (1991). *TIBS* **17**, 241-245.

Ruffle, S. V., Donnelly, D., Blundell, T. and Nugent, J. H. A. (1992). *Photosynth. Res.* **34**, 287-300.

Sopory, S. K., Ghirardi, M. L., Greenberg, B. M., Elich, T., Edelman, M. and Mattoo, A. K. (1993). In: "*Photosynthesis: Photoreactions to Plant Productivity*", edited by Y. P. Abrol, P. Mohanty and Govindjee, Kluwer, Dordrecht, pp. 131-156.

Svensson, B., Vass, I. and Styring, S. (1990). *EMBO J.* **9**, 2051-2059.

Svensson, B., Vaas, I. and Styring, S. (1992). *Z. Naturforsch* **46c**, 765-776.

Swofford, D. L. (1993). PAUP: Phylogenetic analysis using parsimony, version 3.1.1 Computer program distributed by the Illinois Natural History Survey, Champaign, IL.