

Understanding Patent Documents and Data Fields

Paul Oldham (Ph.D.)

WIPO Open Source Patent Analytics Project



What is a Patent?

- A patent can be described in two main ways:
 1. As a particular form of intellectual property right.
 2. As a type of document.

As a form of intellectual property right

1. A patent is a temporary grant of an exclusive right to a patentee to prevent others from making, using, offering for sale, or importing, a patented invention without their consent, in a country where a patent is in force.
2. Patent rights are territorial rights - they are only valid in the territory of the country where granted.
3. Patents are typically granted for a period of 20 years from the filing date of an application but may be opposed or revoked.
4. To be eligible a claimed invention must: a) Involve patentable subject matter; b) Be new or novel; c) Involve an inventive step; d) Be susceptible to industrial application or useful.

Patents as a type of document

- For patent analytics we need to concentrate on patents as documents and to understand:
 1. The structure of patent documents and their data fields.
 2. The strengths and weaknesses of patent databases as a means for obtaining patent data.
- In this session we deal with the basics of patent documents and their data fields.

Basic Data Types

When performing patent analysis we are dealing with data of seven different types:

1. **Dates** (priority, application and publication dates)
2. **Numbers** (priority number, application number, publication number, family members, citations)
3. **Names** (Applicants - also known as Assignees - and Inventors)
4. **Classification codes** (e.g. International Patent Classification/Cooperative Patent Classification)
5. **Text fields** (Title, Abstract, Description, Claims, Sequence data)
6. **Images** (Diagrams)
7. **Additional Information** (Legal Status, Public Registry etc)



Synthetic genomes: The next step for the synthetic genome

Monya Baker

Nature 473, 403–408 (19 May 2011) | doi:10.1038/473403a

Published online 18 May 2011

PDF
 Citation
 Reprints
 Rights & permissions
 Article metrics

Biologists have copied an existing genetic code, but haven't yet commercialized it or written their own. What will it take for a tour de force to reach industrial force?

Subject terms: [Biotechnology](#) • [Genetics and genomics](#)

- Introduction

[Introduction](#) • [References](#) • [Author information](#) • [Related links](#)

A year ago this week, headlines trumpeted that humans had created artificial life. Scientists at the J. Craig Venter Institute in Rockville, Maryland, had chemically synthesized DNA and placed it inside a bacterial cell emptied of its own genetic material. Tests a few days after the insertion showed that the 1-million-base-pair-long synthetic genome was able to run the cellular machinery¹.

Whole-genome engineering could one day create cells unbound by biochemistry as we know it, says George Church, a geneticist at Harvard Medical School in Boston, Massachusetts. Researchers might even be able to design a new genetic code, one that could incorporate more than the 20 or so amino acids used by natural living systems. That achievement is "going to be more than an increment", says Church, "that's going to be a game-changer". But current reality is more prosaic. As

Editors' pick



Image credit: Jessica Fortner

The Anthropocene debate: Momentum is building to establish a new geological epoch recognizing humanity's impact on the planet but there is fierce debate behind the scenes. ►

Science jobs

Science events

[naturejobs.com](#)

Faculty Positions Available in Southwest University
SOUTHWEST UNIVERSITY

Director, Saskatchewan Centre for Cyclotron Sciences
Sylvia Fedoruk Canadian Centre for Nuclear Innovation

NPU Welcomes International Talent to Join Us in 2015
Northwestern Polytechnical University

[Post a free job](#) ► | [More science jobs](#) ►

Most read



(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau(43) International Publication Date
28 February 2008 (28.02.2008)

PCT

(10) International Publication Number
WO 2008/024129 A2

(51) International Patent Classification:

C07H 21/04 (2006.01) C12P 1/04 (2006.01)
C12N 5/06 (2006.01)

(21) International Application Number:

PCT/US2006/046803

(22) International Filing Date:

6 December 2006 (06.12.2006)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

60/742,542 6 December 2005 (06.12.2005) US

(71) Applicant (for all designated States except US): **J. CRAIG VENTER INSTITUTE** [US/US]; 9704 Medical Center Drive, Rockville, MD 20850 (US).(71) Applicant (for US only): **HUTCHISON, Clyde, A., III** [US/US]; c/o J. Craig Venter Institute, 9704 Medical Center Drive, Rockville, MD 20850 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **VENTER, Craig, J.** [US/US]; c/o J. Craig Venter Institute, 9704 Medical Center Drive, Rockville, MD 20850 (US). **SMITH, Hamilton, O.** [US/US]; c/o J. Craig Venter Institute, 9704 Medical Center Drive, Rockville, MD 20850 (US).(74) Agents: **BATHURST, Brian** et al.; Carr & Ferrell LLP, 2200 Geng Road, Palo Alto, CA 94303 (US).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LV, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, SV, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH,

[Continued on next page]

(54) Title: SYNTHETIC GENOMES



(57) Abstract: Methods are provided for constructing a synthetic genome, comprising generating and assembling nucleic acid cassettes comprising portions of the genome, wherein at least one of the nucleic acid cassettes is constructed from nucleic acid components that have been chemically synthesized, or from copies of the chemically synthesized nucleic acid components.

[Refine search](#) → [Results](#) → [IL192041 \(A\)](#) → [Family](#) → ... → [Family](#) → [WO2008024129 \(A2\)](#)

WO2008024129 (A2)

Bibliographic data

Description

Claims

Mosaics

Original document

Cited documents

Citing documents

INPADOC legal status

INPADOC patent family

Bibliographic data: WO2008024129 (A2) — 2008-02-28

★ In my patents list

Previous ◀ 9 / 9

Next ▶ EP Register

Report data error

Print

SYNTHETIC GENOMES

Page bookmark [WO2008024129 \(A2\) - SYNTHETIC GENOMES](#)

Inventor(s): VENTER CRAIG J [US]; SMITH HAMILTON O [US] ±

Applicant(s): CRAIG VENTER INST J [US]; HUTCHISON CLYDE A III [US]; VENTER CRAIG J [US]; SMITH HAMILTON O [US] ±

Classification:
- international: **C07H21/04; C12N5/06; C12P1/04**
- cooperative: **C12N15/10; C12N15/1093; C12N15/66**

Application number: WO2006US46803 20061206

Priority number(s): [US20050742542P 20051206](#)

Also published as: [WO2008024129 \(A3\)](#) [US2007264688 \(A1\)](#) [JP2009518038 \(A\)](#) [JP5106412 \(B2\)](#) [IL192041 \(A\)](#)
→ [more](#)

Abstract of WO2008024129 (A2)

Translate this text into ⓘ

Albanian

 powered by EPD and Google

Methods are provided for constructing a synthetic genome, comprising generating and assembling nucleic acid cassettes comprising portions of the genome, wherein at least one of the nucleic acid cassettes is constructed from nucleic acid components that have been chemically synthesized, or from copies of the chemically synthesized nucleic acid components. In one embodiment, the entire synthetic genome is constructed from nucleic acid components that have been chemically synthesized, or from copies of the chemically synthesized nucleic acid components. Rational methods may be used to design the synthetic genome (e.g., to establish a minimal genome and/or to optimize the function of genes within a genome, such as by mutating or rearranging the order of the genes).; Synthetic genomes of the invention may be introduced into vesicles (e.g., bacterial cells from which part or all of the resident genome has been removed, or synthetic vesicles) to generate synthetic cells. Synthetic genomes or synthetic cells may be used for a variety of purposes, including the generation of synthetic fuels, such as hydrogen or ethanol.



Quick help

- [What is meant by high quality text as facsimile?](#)
- [What does A1, A2, A3 and B stand for after a European publication number?](#)
- [What happens if I click on "In my patents list"?](#)
- [What happens if I click on the "Register" button?](#)
- [Why are some sidebar options deactivated for certain documents?](#)
- [How can I bookmark this page?](#)
- [Why does a list of documents with the heading "Also published as" sometimes appear, and what are these documents?](#)
- [What is Global dossier?](#)
- [Why do I sometimes find the abstract of a corresponding document?](#)
- [What happens if I click on the red "patent translate" button?](#)

◀ About Espacenet Other EPO online services ▼

Search

Result list

★ My patents list (0)

Query history

Settings

Help

Refine search → Results → [IL192041 \(A\)](#) → Family → ... → Family → [WO2008024129 \(A2\)](#)

WO2008024129 (A2)

Bibliographic data

Description

Claims

Mosaics

Original document

Cited documents

Citing documents

INPADOC legal status

INPADOC patent family

Description: WO2008024129 (A2) — 2008-02-28

★ In my patents list

Previous ◀

9 / 9

Next ▶

EP Register

Report data error

Print

SYNTHETIC GENOMES

Description of WO2008024129 (A2)

A high quality text as facsimile in your desired language may be available amongst the following family members:

[AU2006347573 \(B2\)](#)
[CA2643356 \(A1\)](#)
[CN101501207 \(A\)](#)
[EP1968994 \(B1\)](#)
[JP5106412 \(B2\)](#)
[US2007264688 \(A1\)](#)

Translate this text into ⓘ

Albanian



patenttranslate

powered by EPO and Google

The EPO does not accept any responsibility for the accuracy of data and information originating from other authorities than the EPO; in particular, the EPO does not guarantee that they are complete, up-to-date or fit for specific purposes.

SYNTHETIC GENOMES

By J. Craig Venter, Hamilton O. Smith and Clyde A. Hutchison III

CROSS-REFERENCE TO RELATED APPLICATIONS

[001] The present application claims benefit and priority from U.S. Provisional Patent Application Serial No. 60/742,542 filed on Dec. 6, 2005, entitled, "Synthetic Genomes;" the present application is related to U.S. Provisional Patent Application Serial No. 60/752,965 filed on Dec. 23, 2005, entitled, "Introduction of Genomes into Microorganisms;" U.S. Provisional Patent Application Serial No. 60/741,469 filed on Dec. 2, 2005, entitled, "Error Correction Method;" and U.S. Non-Provisional Patent Application Serial No. 11/502,746 filled on Aug. 11, 2006, entitled "In Vitro Recombination Method," all of which are incorporated herein by reference.

STATEMENT REGARDING FEDERALLY SPONSORED RESEARCH OR DEVELOPMENT

[002] This invention was made with U.S. government support (DOE grant number DE-FG02-02ER63453). The government has certain rights in the invention.

Quick help

- [What is meant by high quality text as facsimile?](#)
- [What happens if I click on "In my patents list"?](#)
- [What happens if I click on the "Register" button?](#)
- [What happens if I click on the red "patent translate" button?](#)
- [Why is the description sometimes in French or German or another language altogether?](#)
- [How can I search in the text of the description?](#)
- [What is Global dossier?](#)
- [How can I view chemical structures in the full text?](#)

◀ About Espacenet Other EPO online services ▶

Search Result list ★ My patents list (0) Query history Settings Help

Refine search → Results → IL192041 (A) → Family → ... → Family → WO2008024129 (A2)

WO2008024129 (A2)

Bibliographic data

Description

Claims

Mosaics

Original document

Cited documents

Citing documents

INPADOC legal status

INPADOC patent family

Claims: WO2008024129 (A2) — 2008-02-28

★ In my patents list Previous 9 / 9 Next EP Register Report data error Print

SYNTHETIC GENOMES

Claims of WO2008024129 (A2)

A high quality text as facsimile in your desired language may be available amongst the following family members:

▢ AU2006347573 (B2) ▢ CA2643356 (A1) ▢ CN101501207 (A) ▢ EP1968994 (B1) ▢ JP2009518038 (A) ▢ US2007264688 (A1)

Translate this text into ⓘ

Albanian

patenttranslate powered by EPO and Google

Quick help

- [What is meant by high quality text as facsimile?](#)
- [What happens if I click on "In my patents list"?](#)
- [What happens if I click on the "Register" button?](#)
- [What happens if I click on the red "patent translate" button?](#)
- [How can I view the claim structure?](#)
- [Why are the claims sometimes in French or German or another language altogether?](#)
- [How can I search in the text of the claims?](#)
- [What is Global dossier?](#)
- [How can I view chemical structures in the full text?](#)

Original claims

Claims tree

The EPO does not accept any responsibility for the accuracy of data and information originating from other authorities than the EPO; in particular, the EPO does not guarantee that they are complete, up-to-date or fit for specific purposes.

CLAIMS What is claimed is:

1. A method for constructing a synthetic genome comprising: assembling nucleic acid cassettes that comprise portions of the synthetic genome, wherein at least one of the nucleic acid cassettes is constructed from nucleic acid components that have been chemically synthesized, or from copies of chemically synthesized nucleic acid components.
2. The method of claim 1, wherein one or more of the nucleic acid cassettes are prepared by assembling chemically synthesized, overlapping oligonucleotides of about 50 nucleotides.
3. The method of claim 1, wherein the cassettes are about 4 kilobases to about 7 kilobases in length.
4. The method of claim 1, wherein the cassettes are about 4.5 kilobases to about 6.5 kilobases in length.



◀ About Espacenet Other EPO online services ▼

Search

Result list

★ My patents list (0)

Query history

Settings

Help

Refine search → Results → IL192041 (A) → Family → ... → WO2008024129 (A2) → Family

WO2008024129 (A2)

Bibliographic data

Description

Claims

Mosaics

Original document

Cited documents

Citing documents

INPADOC legal status

INPADOC patent family

Quick help

- [Can I export this list?](#)
- [What happens if I click on "Download covers"?](#)
- [Can I sort the list?](#)
- [What happens if I click on the star icon?](#)
- [What is a patent family?](#)
- [What happens if I tick the "show citations" box?](#)
- [What is an INPADOC patent family?](#)
- [Are all the documents in an INPADOC family equivalents?](#)
- [Why is the same document published several times in the same country?](#)

Family list: WO2008024129 (A2) — 2008-02-28

☐ Select all (0/9) ☐ Compact ☐ Export (CSV | XLS) ☐ Download covers ☐ CCD

9 application(s) for: WO2008024129 (A2)

Sort by Sort order ☐ show citations

☐ 1. SYNTHETIC GENOMES

★ Inventor:	Applicant:	CPC:	IPC:	Publication info:	Priority date:
VENTER CRAIG J [US] SMITH HAMILTON O [US]	CRAIG VENTER INST J [US] HUTCHISON CLYDE A III [US] (+2)	C12N15/10 C12N15/1093 C12N15/66	C07H21/04 C12N5/06 C12P1/04	WO2008024129 (A2) 2008-02-28 WO2008024129 (A3) 2008-10-09	2005-12-06

☐ 2. Synthetic genomes

★ Inventor:	Applicant:	CPC:	IPC:	Publication info:	Priority date:
SMITH HAMILTON O VENTER CRAIG J	CRAIG VENTER INST J	C12N15/10 C12N15/1093 C12N15/66	C07H21/04 C12N5/06 C12P1/04	AU2006347573 (A1) 2008-02-28 AU2006347573 (B2) 2013-01-17	2005-12-06

☐ 3. SYNTHETIC GENOMES

★ Inventor:	Applicant:	CPC:	IPC:	Publication info:	Priority date:
HUTCHISON CLYDE A III [US] SMITH HAMILTON O [US] (+1)	CRAIG VENTER INST J [US]	C12N15/10 C12N15/1093 C12N15/66	C07H21/00 C07H21/04 C12N1/00 (+4)	CA2643356 (A1) 2008-02-28	2005-12-06

☐ 4. Synthetic genomes

★ Inventor:	Applicant:	CPC:	IPC:	Publication info:	Priority date:
VENTER CRAIG J	CRAIG VENTER INST J [US]	C12N15/10	C07H21/00	CN101501207 (A)	2005-12-06





◀ About Espacenet Other EPO online services ▼

Search Result list My patents list (0) Query history Settings Help

Refine search → Results → IL192041 (A) → Family → ... → WO2008024129 (A2) → Citations

WO2008024129 (A2)

Bibliographic data

Description

Claims

Mosaics

Original document

Cited documents

Citing documents

INPADOC legal status

INPADOC patent family

Quick help —

- [What are cited documents?](#)
- [Can I export this list?](#)
- [What happens if I click on "Download covers"?](#)
- [What happens if I click on the star icon?](#)

Cited documents: WO2008024129 (A2) — 2008-02-28

☐ Select all (0/4) Compact Export (CSV | XLS) Download covers Print

4 documents cited in relation to WO2008024129 (A2)

Sort by Sort order

International search citation

☐ 1. [Method for the complete chemical synthesis and assembly of genes and genomes](#)

★	Inventor: EVANS GLEN A [US]	Applicant: EGEA BIOSCIENCES INC [US]	CPC: B01J19/0046 B01J2219/00317 B01J2219/00511 (+17)	IPC: B01J19/00 C12N15/10 C12N15/66 (+7)	Publication info: US6521427 (B1) 2003-02-18	Priority date: 1997-09-16
---	---	---	---	--	--	-------------------------------------

International search NPL citation


☐ 2. [Venter aims for maximum impact with minimal genome](#)

★	Author: Erika Check	Publication data: NATURE, 20021128 Nature Publishing Group, United Kingdom	CPC:	Source information: Vol:420,Nr:6914,Page(s):350	Publication info: XP008128716
---	-------------------------------	--	-------------	---	---

☐ 3. [Generating a synthetic genome by whole genome assembly: phiX174 bacteriophage from synthetic oligonucleotides](#)

★	Author: SMITH HAMILTON O ET AL	Publication data: Proceedings of the National Academy of Sciences, 20031223 National Academy of Sciences, US	CPC:	Source information: Vol:100,Nr:26,Page(s):15440 - 15445	Publication info: XP002301506
---	---	---	-------------	--	---

◀ About Espacenet Other EPO online services ▼

Search Result list  My patents list (0) Query history Settings Help

Refine search → Results → IL192041 (A) → Family → ... → WO2008024129 (A2) → Citations

WO2008024129 (A2)

Bibliographic data

Description

Claims

Mosaics

Original document

Cited documents

Citing documents

INPADOC legal status

INPADOC patent family

Quick help —

- [Can I export this list?](#)
- [What happens if I click on "Download covers"?](#)
- [What are citing documents?](#)
- [Why do some documents not have any citing documents?](#)
- [What happens if I click on the star icon?](#)

Citing documents: WO2008024129 (A2) — 2008-02-28

☐ Select all (0/2)  Compact  Export (CSV | XLS)  Download covers

 Print

2 documents citing WO2008024129 (A2)

Sort by Sort order

☐ 1. BACTERIAL ENGINEERING

★	Inventor: WILLIAMS DAVID HUGH [GB] TURNER ARTHUR KEITH [GB] (+1)	Applicant: DISCUVA LTD [GB]	CPC: C12N15/102 C12N15/1082	IPC: C12N15/10	Publication info: WO2014072697 (A1) 2014-05-15	Priority date: 2012-11-06
---	--	---------------------------------------	--	--------------------------	---	-------------------------------------

☐ 2. METHODS FOR CLONING AND MANIPULATING GENOMES

★	Inventor: BENDERS GWYNEDD A [US] GLASS JOHN I [US] (+10)	Applicant: SYNTHETIC GENOMICS INC [US] BENDERS GWYNEDD A [US] (+11)	CPC: C12N15/1031 C12N15/1079 C12N15/66 (+1)	IPC: C12N15/10 C12N15/74	Publication info: WO2011109031 (A1) 2011-09-09 WO2011109031 (A8) 2012-09-20	Priority date: 2010-03-05
---	---	--	--	---------------------------------------	--	-------------------------------------



◀ About Espacenet Other EPO online services ▼

Search

Result list

★ My patents list (0)

Query history

Settings

Help

WO2008024129 (A2)

Bibliographic data

Description

Claims

Mosaics

Original document

Cited documents

Citing documents

INPADOC legal status

INPADOC patent family

Quick help —

- [What happens if I click on "In my patents list"?](#)
- [What happens if I click on the "Register" button?](#)
- [What does "legal status" mean?](#)
- [Why is the legal status not always available?](#)
- [How might this information be useful to me?](#)
- [How reliable is this data?](#)

INPADOC legal status: WO2008024129 (A2) — 2008-02-28

★ In my patents list → EP Register 📄 Report data error

🖨️ Print

SYNTHETIC GENOMES

The EPO does not accept any responsibility for the accuracy of data and information originating from other authorities than the EPO; in particular, the EPO does not guarantee that they are complete, up-to-date or fit for specific purposes.

Legal status of WO2008024129 (A2) 2008-02-28; WO2008024129 (A3) 2008-10-09:

WO	F	2006046803 W (Patent of invention)
Event date :	2008/05/07	
Event code :	121	
Code Expl.:	EP: THE EPO HAS BEEN INFORMED BY WIPO THAT EP WAS DESIGNATED IN THIS APPLICATION	
CC OF CORRESP. PAT. :	EP	
CORRESP. PATENT D. :	06851474	
KD OF CORRESP. PAT. :	A2	
Event date :	2008/06/06	
Event code :	WWE	
Code Expl.:	+ WIPO INFORMATION: ENTRY INTO NATIONAL PHASE	
CC OF CORRESP. PAT. :	JP	
CORRESP. PATENT D. :	2008544524	
Event date :	2008/06/07	
Event code :	NENP DE	



Round Up

In this session we have walked through some of the most important patent data fields.

These fields are the building blocks for sophisticated patent analysis. In future sessions we will focus on:

- Retrieving data with these fields
- Cleaning up the data in these fields
- Mapping trends
- Network Mapping
- Geographic Mapping

Learn More

- Visit the Git hub project home page.
- Access the materials direct on Github here and add your own.
- Email me with comments, corrections or suggestions.