nature Vol 466|26 August 2010|doi:10.1038/nature09294

LETTERS

Direct formation of supermassive black holes via multi-scale gas inflows in galaxy mergers

L. Mayer[1], S. Kazantzidis[2,3], A. Escala[4,5] & S. Callegari[1]

Observations of distant quasars indicate that supermassive black holes of billions of solar masses already existed less than a billion years after the Big Bang[1]. Models in which the 'seeds' of such black holes form by the collapse of primordial metal-free stars[2,3] cannot explain the rapid appearance of these supermassive black holes because gas accretion is not sufficiently efficient[4–6]. Alternatively, these black holes may form by direct collapse of gas within isolated protogalaxies[7,8], but current models require idealized conditions, such as metal-free gas, to prevent cooling and star formation from consuming the gas reservoir[9–11]. Here we report simulations showing that mergers between massive protogalaxies naturally produce the conditions for direct collapse into a supermassive black hole with no need to suppress cooling and star formation. Merger-driven gas inflows give rise to an unstable, massive nuclear gas disk of a few billion solar masses, which funnels more than 10^8 solar masses of gas to a sub-parsec-scale gas cloud in only 100,000 years. The cloud undergoes gravitational collapse, which eventually leads to the formation of a massive black hole. The black hole can subsequently grow to a billion solar masses on timescales of about 10^8 years by accreting gas from the surrounding disk.

To grow into the billion-solar-mass supermassive black holes that power bright quasars at redshift $z \approx 6$, the hundred-solar-mass black-hole seeds produced by the collapse of primordial metal-free (population III) stars at $z > 15$ (refs 2, 9, 10) need to accrete steadily at or above the Eddington rate[3]. However, such high accretion rates cannot be maintained because the ionized gas surrounding the seeds has very low densities[4], with radiative feedback from accretion and radiation pressure exacerbating the problem of the inefficient gas capture[5,6,12]. Dynamical effects, such as the expulsion of a black hole from the host halo owing to asymmetric emission of gravitational waves in mergers (the "gravitational rocket") can further stifle the growth of the seeds[3,13]. In the alternative model, direct collapse[7–10,14], efficient outward transport of angular momentum by gravitational torques[15–17], gravito-driven turbulence[18], or magneto-rotational instability[19] in isolated protogalactic disks[8,10,16] or primordial halos[11,20] can in principle accumulate enough gas at their centres to produce seeds exceeding 10^5 solar masses ($M_\odot$). Such massive seeds should be able to grow rapidly to a billion solar masses[3]. However, gas would be converted into stars faster than it is driven to the centre[16,20]. Hence it is necessary to invoke primordial metal-free gas and dissociation of the star-forming molecular gas phase by the cosmic ultraviolet background in order to suppress cooling and star formation[8,10,16,21]. Yet recent calculations show that the metals readily produced by the early generations of stars are sufficient to trigger efficient cooling and star formation[9]. Therefore, whether or not protogalaxies ever met the conditions required for direct supermassive black-hole formation is still a puzzle.

Until now, direct collapse scenarios were limited to considering gas inflows in isolated objects. On the contrary, in the accepted picture of hierarchical structure formation, protogalaxies grow primarily through rapid and repeated mutual interactions and mergers[21,22]. In mergers between galaxies of comparable mass—the so-called "major mergers"—tidal torques and shock dissipation are capable of driving most of the gas content of galaxies from kiloparsec scales down to scales of several tens of parsecs at rates as high as 100 $M_\odot$ per year despite the concurrent high star-formation rates[23,24]. If such prodigious gas inflows could continue all the way down to the very centre of the merger remnant they could provide the natural conditions for direct supermassive black-hole formation. Addressing this issue requires a three-dimensional simulation following gas dynamics across an unprecedented range of spatial scales, from tens of kiloparsecs to below a parsec, which has been computationally too demanding until now.

We begin by performing a merger simulation between two identical disk galaxies with moderate amounts of gas, several kiloparsecs in size and embedded in 10^{12} $M_\odot$ cold-dark-matter halos (see Supplementary Information). The mass of the dark halos is consistent with that inferred for the hosts of high-redshift quasars on the basis of their number densities[3,5,22]. The two galaxies are placed on a parabolic orbit whose parameters are consistent with those found in cosmological simulations[23,24]. We employ the technique of smoothed particle hydrodynamics, modelling the galaxy collision with radiative cooling, star formation and a temperature floor at 20,000 K to mimic non-thermal pressure due to turbulence in the interstellar medium (see Supplementary Information).

The two merging galaxies undergo two close encounters as dynamical friction against their extended halos erodes their orbital energy. When the two baryonic cores are separated by six kiloparsecs and begin their last orbit before the final impact we perform particle splitting in the gas component within a volume 30 kiloparsecs in size (see Supplementary Information). As a result, we increase our gas mass resolution eightfold and simultaneously decrease the gravitational softening, achieving a spatial resolution of a tenth of a parsec (see Supplementary Information). In this refined calculation we adopt an effective equation of state that accounts for the net balance between cooling and heating[25]. A lower-resolution unrefined simulation shows that during the final collision the star formation peaks at about 30 $M_\odot$ per year over about 10^8 years (ref. 23). Motivated by this, we chose an equation of state with an effective adiabatic index γ in the range 1.1–1.4, which is appropriate for interstellar gas heated by a major starburst[26], with solar metallicity as suggested by the abundance analysis of the high-redshift quasars [27] (see Supplementary Information).

The final collision of the two galactic cores produces a massive turbulent rotating nuclear disk with a mass of about two billion $M_\odot$ and a radius of about 80 pc. The disk is born in an unstable configuration, with a prominent two-armed spiral pattern imprinted by the collision and sustained by its own strong self-gravity (Fig. 1). The gas has a high turbulent velocity ($\sim$100 km s^{-1}) maintained by gravitational

---

[1]Institute for Theoretical Physics, University of Zürich, 190 Winterthurestrasse, Zürich 8057, Switzerland. [2]Center for Cosmology and Astro-Particle Physics, The Ohio State University, Ohio 43210, USA. [3]Departments of Physics and Astronomy, The Ohio State University, Ohio 43210, USA. [4]Departamento de Astronomia, Universidad de Chile, Casilla 36-D, Santiago 7550000, Chile. [5]Kavli Institute for Particle Astrophysics and Cosmology (KIPAC), Stanford University, 2575 Sand Hill Road MS 29 Menlo Park, California 94025, USA.

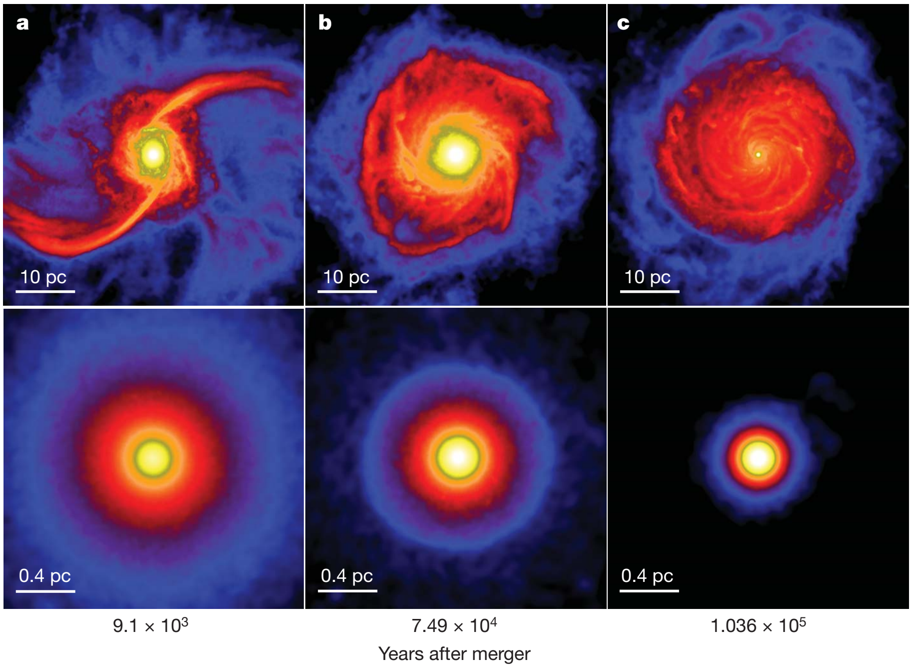

Figure 1 | Time evolution of the nuclear gas disk from its formation until the onset of central collapse. The surface density maps of the nuclear disk are shown at large scales (**a–c**, top) and small scales (**a–c**, bottom). Particles are colour-coded on a logarithmic scale with brighter colours in regions of higher density. The density ranges from 2×10^4 to $10^8 M_{\odot}$ per square parsec (upper panels) and from 2×10^6 to $2 \times 10^{10} M_{\odot}$ per square parsec (lower panels). The time of the merger is defined as the time at which the two density peaks associated with the merging galactic cores are no longer distinguishable. For reference, the disk orbital time at ~20 pc is 5×10^4 years, while at 1 pc it is $\sim 4 \times 10^3$ years. Global spiral modes, in particular the two-armed spiral initially triggered by the final collision between the two cores, are evident at scales of tens of parsecs (top panels of **a** and **b**) and cause the mass increase in the central parsec region (bottom panels) that allows the collapse into a massive central cloud (bottom panels of **b** and **c**).

instability[18] and rotates at a speed of several hundred kilometres per second within 50 pc. The disk is stable against fragmentation (see Supplementary Fig. 4) but the strong spiral pattern swiftly transports mass inward and angular momentum outward (see Supplementary Information). About 10,000 years after the merger is completed more than 20% of the disk mass (at least $5 \times 10^8 M_{\odot}$) has already accumulated within the central few parsecs (Fig. 2) where the inflow rate peaks at over $10^4 M_{\odot}$ per year (see Supplementary Fig. 3), orders of magnitude above inflow rates in isolated protogalaxies[10,16,17,20].

The gas funnelled to the central region (2–3 parsecs) of the nuclear disk settles into a rotating, pressure-supported cloud. The density of the cloud continuously increases as it gains mass from the inflow until it becomes Jeans unstable and collapses to sub-parsec scales on the local dynamical time of 1,000 years (Fig. 1). The supermassive cloud contains around 13% of the disk mass ($\sim 2.6 \times 10^8 M_{\odot}$) (Fig. 2). The simulation is stopped once the central cloud has contracted to a size comparable to the spatial resolution limit. At this point the cloud is still Jeans unstable. With greater resolution its collapse should continue because the equation of state would become essentially isothermal at even higher densities[27]. With its steep density profile ($\rho \approx r^{-\alpha}$, with $\alpha > 2$, where r is the radius of the cloud and ρ is its density), at radii as small as a thousandth of a parsec the cloud would be as massive and dense as a quasi-star that can collapse directly into a massive black hole[9]. Alternatively, the cloud could collapse into a supermassive star on its free-fall timescale, which would later produce a black hole of comparable mass[20].

The formation of the collapsing cloud occurs in just 100,000 years after the completion of the merger, a timescale much shorter than the 10^8 years needed to convert most of the nuclear gas into stars during the starburst (see Supplementary Information). Therefore the merger-driven gas collapse overcomes the major difficulty of previous direct collapse models in isolated protogalaxies: such models were forced to suppress star formation owing to much weaker inflows[10,16,20]. The rapid accumulation of more than $10^8 M_{\odot}$ of gas in the central parsec is confirmed by a complementary simulation which incorporates star formation (see Supplementary Information and Supplementary Figs 1 and 2).

Assuming that the nuclear disk forms stars with an efficiency of about 30%, the upper limit deduced from observation and models of star-forming molecular clouds[26], a gas mass in excess of a billion $M_{\odot}$ is available to feed the black hole despite the concurrent starburst. Considering Eddington-limited accretion[28] a seed black hole weighing only $10^5 M_{\odot}$, and hence forming out of less than 1% of the supermassive cloud, can grow to a billion $M_{\odot}$ in as little as 3.6×10^8 years (see Supplementary Information), explaining the rapid appearance of bright quasars if the merger takes place at $z \approx 7$–8. In the concordance WMAP5 cosmology, $10^{12} M_{\odot}$ galaxies would correspond to fairly high-density peaks collapsing at $z \approx 7$–8 (ref. 5), 4 to 5σ, where σ is the root-mean-square variation of the cosmological density field). At $z > 6$ at least one major merger between galaxies this massive is expected[22]. At such early epochs the galaxy merger would take place on a timescale much shorter than a billion years owing to the higher densities and shorter orbital timescales of cold-dark-matter halos compared to the present time[29] (see Supplementary Information). Finally, kiloparsec-scale rotating disks of stars and gas such as those

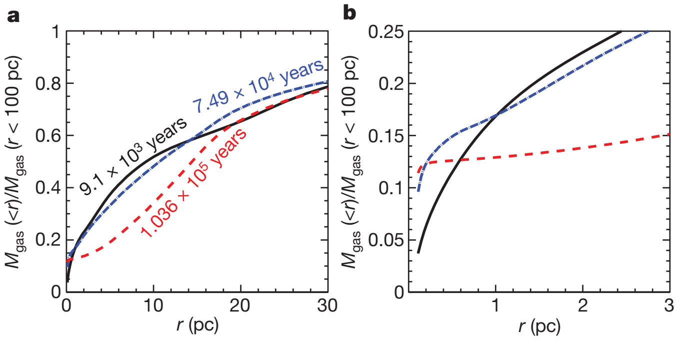

Figure 2 | Evolution of the mass distribution of the nuclear region. a, The cumulative gas mass profile at scales of tens of parsecs is shown. **b**, The cumulative gas mass profile within the inner few parsecs is displayed. Profiles are normalized to the total gas mass within the radius of the nuclear disk (100 pc). The disk radius is determined from the sharp drop in the nuclear gas density distribution. The black, red and blue curves in each panel correspond to the mass profile at the indicated times after the merger, these being the same snapshots used in Fig. 1. Mass redistribution occurs as spiral arms (at scales of tens of parsecs) push mass inward and shed angular momentum outwards, gradually leading to an increasing mass concentration in the very central region (blue line in **b**) and, correspondingly, a flattening of the mass distribution at scales of tens of parsecs (blue line in **a**). This triggers the Jeans collapse of the central inner few parsecs into a supercloud containing ~13% of the total disk mass, which manifests as a strong flattening of the profile at parsec scales as the cloud absorbs a large fraction of the mass in that region (red line in **b**).

that we have assumed in our initial conditions should already be present in halos with masses exceeding $10^{11}M_{\odot}$ at $z > 4$ (ref. 17).

Additional numerical experiments show that galaxies with masses a few times $10^{11}M_{\odot}$ can develop an inflow even stronger than that in our reference simulation with $10^{12}M_{\odot}$ galaxies (see Supplementary Information and Supplementary Figs 2 and 5). This can be understood in terms of the factors that determine the spiral instability in the nuclear disk, namely its density and temperature. At a few times $10^{11}M_{\odot}$ the gaseous components of the galaxies are still prominent enough to produce a dense self-gravitating nuclear disk after the galaxy merger, but such a disk is colder and thus even more unstable to spiral modes than is our reference simulation. The shock heating during the final collision between the two galaxies is lower because the orbital energy dissipated into thermal energy by the shock decreases as the mass of the merging galaxies decreases (see Supplementary Information). Therefore our direct collapse mechanism could have produced massive black-hole seeds in galaxies that were more common than the high-σ peaks hosting the bright quasars at $z > 6$, including the most massive progenitors of our own Milky Way (see Supplementary Information). On the contrary, we find that in merging galaxies with masses well below $10^{11}M_{\odot}$ the inflow is too weak to trigger the gravitational collapse of the central region. The light nuclear disk that forms has a very low gas surface density that cannot support a strong spiral instability despite the fact that shock heating at the final impact is also low. This is consistent with the absence of massive black holes in dwarf galaxies.

The exact galaxy mass threshold between the two regimes of unstable and stable nuclear disks will depend on the detailed balance between radiative cooling and heating in the gas, and on the competing effect of gas outflows from supernovae explosions. Recent results showing that in galaxies with masses of a few tens of billions of $M_{\odot}$ supernovae-driven gas outflows prevail over merger-driven gas inflows[30] strengthen our conclusion that formation of massive black-hole seeds by direct collapse would not occur at low galaxy masses.

Large deviations from the local $M_{\rm BH}$–$\sigma_{\rm los}$ relation[21] (where $M_{\rm BH}$ is the mass of the black hole and $\sigma_{\rm los}$ is the line-of-sight stellar velocity dispersion of the central stellar spheroid) should be expected at high redshift as the black-hole mass grows faster than that of its host galaxy in our scenario. At low redshift our formation mechanism should be suppressed as pre-existing supermassive black-holes in the two merging galaxies release energy into the surrounding medium, preventing instabilities and inflows in the nuclear disk, while light seed black holes formed at higher redshift from the collapse of a population III star would not hamper direct collapse owing to their modest energetic feedback on the gas (see Supplementary Information).

Future gravitational wave experiments such as the Laser Interferometer Space Antenna (LISA) will be able to test for the existence of a population of black holes that had a 'jump start' by probing the mass distribution of merging black holes as a function of redshift.

Received 26 April; accepted 21 June 2010.

1. Fan, X. Evolution of high redshift quasars. *New Astron. Rev.* **50,** 665–671 (2006).
2. Madau, P. & Rees, M. J. Massive black holes as population III remnants. *Astrophys. J.* **551,** L27–L30 (2001).
3. Volonteri, M. & Rees, M. J. Quasars at z=6: the survival of the fittest. *Astrophys. J.* **650,** 669–678 (2006).
4. Johnson, J. L. & Bromm, V. The aftermath of the first stars: massive black holes. *Mon. Not. R. Astron. Soc.* **474,** 1557–1568 (2007).
5. Pelupessy, F. I., di Matteo, T. & Ciardi, B. How rapidly do supermassive black hole "seeds" grow at early times? *Astrophys. J.* **665,** 107–119 (2007).
6. Alvarez, M. A., Wise, J. H. & Abel, T. Accretion onto the first stellar mass black holes. *Astrophys. J.* **701,** L133–L137 (2009).
7. Begelman, M., Volonteri, M. & Rees, M. J. Formation of supermassive black holes by direct collapse in pre-galactic halos. *Mon. Not. R. Astron. Soc.* **370,** 289–298 (2006).
8. Volonteri, M., Lodato, G., &. Natarayan, P. The evolution of massive black hole seeds. *Mon. Not. R. Astron. Soc.* **383,** 1079–1088 (2008).
9. Omukai, K., Schneider, R. & Haiman, Z. Can supermassive black holes form in metal-enriched high redshift protogalaxies? *Astrophys. J.* **686,** 801–814 (2008).
10. Regan, J. & Haehnelt, M. Pathways to massive black holes and compact star clusters in pre-galactic dark matter haloes with virial temperatures >10000 K. *Mon. Not. R. Astron. Soc.* **396,** 343–353 (2009).
11. Bromm, V. & Loeb, A. Formation of the first supermassive black holes. *Astrophys. J.* **596,** 34–36 (2003).
12. Milosavljevic, M., Couch, S. M. & Bromm, V. Accretion onto intermediate-mass black holes in dense protogalactic clouds. *Astrophys. J.* **696,** L146–L149 (2009).
13. Haiman, Z. Constraints from gravitational recoil on the growth of supermassive black holes at high redshift. *Astrophys. J.* **613,** 36–40 (2004).
14. Haehnelt, M. G., Natarayan, P. & Rees, M. J. High-redshift galaxies, their active nuclei and central black holes. *Mon. Not. R. Astron. Soc.* **300,** 817–827 (1998).
15. Shlosman, J., Frenk, J. & Begelman, M. C. Bars within bars—a mechanism for fueling galactic nuclei. *Nature* **338,** 45–47 (1989).
16. Lodato, G. & Natarayan, P. Supermassive black hole formation during the assembly of pregalactic discs. *Mon. Not. R. Astron. Soc.* **371,** 1813–1823 (2006).
17. Levine, R., Gnedin, N., Hamilton, A. J. S., &. Kravtsov, A. Resolving gas dynamics in the circumnuclear region of a disk galaxy in a cosmological simulation. *Astrophys. J.* **678,** 154–167 (2008).
18. Escala, A. Toward a comprehensive fueling-controlled theory of the growth of massive black holes and host spheroids. *Astrophys. J.* **671,** 1264–1271 (2007).
19. Eisenstein, D. & Loeb, A. Origin of quasar progenitors from the collapse of low-spin cosmological perturbations. *Astrophys. J.* **443,** 11–17 (1995).
20. Wise, J., Turk, M. & Abel, T. Resolving the formation of protogalaxies. II. Central gravitational collapse. *Astrophys. J.* **682,** 745–757 (2008).
21. Djorgovski, S. G., Volonteri, M., Springel, V., Bromm, V. & Meylan, G. The origins and the early evolution of quasars and supermassive black holes. In *Proceedings of the XIth Marcel Grossmann Meeting on General Relativity* (eds Kleinert, H. Jantzen R.T. & Ruffini, R.) (World Scientific, in the press).
22. Li, Y. *et al.* Formation of $z \sim 6$ quasars from hierarchical mergers. *Astrophys. J.* **665,** 187–208 (2007).
23. Kazantzidis, S. *et al.* The fate of supermassive black holes and the evolution of the MBH-relation in merging galaxies: the effect of gaseous dissipation. *Astrophys. J.* **623,** L67–L70 (2005).
24. Mayer, L. *et al.* Rapid formation of supermassive black hole binaries in galaxy mergers with gas. *Science* **316,** 1874–1877 (2007).
25. Spaans, M. & Silk, J. The polytropic equation of state of interstellar gas clouds. *Astrophys. J.* **538,** 115–120 (2000).
26. Klessen, R. S., Spaans, M. & Jappsen, A. The stellar mass spectrum in warm and dusty gas: deviations from Salpeter in the Galactic centre and in circumnuclear starburst regions. *Mon. Not. R. Astron. Soc.* **374,** L29–L33 (2007).
27. Walter, F. *et al.* Resolved molecular gas in a quasar host galaxy at redshift z=6.42. *Astrophys. J.* **615,** L17–L20 (2004).
28. Shapiro, S. L. Spin, accretion, and the cosmological growth of supermassive black holes. *Astrophys. J.* **620** 59–68 (2005).
29. Mo, H., Mao, S. & White, S. D. M. The formation of galactic discs. *Astrophys. J.* **295,** 319–336 (1998).
30. Governato, F. *et al.* Bulgeless dwarf galaxies and dark matter cores from supernova-driven outflows. *Nature* **463,** 203–206 (2010).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank M. Colpi, R. Durisen, F. Governato, P. Madau, T. Quinn, M. Volonteri, D. Weinberg, R. Teyssier and S. White for discussions. This research has been supported by the Swiss National Science Foundation (SNF), by the Center for Cosmology and Astro Particle Physics (CCAPP) at Ohio State University, and by the Kavli Institute for Particle Astrophysics (KIPAC) at Stanford University. L.M., S.K. and S.C. acknowledge the Kavli Institute for Theoretical Physics at the University of California in Santa Barbara (KITP) for hospitality during the initial stages of this work on the "Building the Milky Way" programme. L.M. also thanks the KITP for hospitality during the final completion of the article while on the "Exoplanets Rising" programme. All computations were performed on the Zbox3 supercomputer at the University of Zürich and on the Brutus cluster at ETH Zürich.

Author Contributions L.M. provided the scientific leadership, led the design of the numerical experiments, wrote the paper and led the analysis and interpretation of the simulations. S.K. and A.E. contributed to the initial concept of the numerical experiments and to their analysis and interpretation. They also helped with the writing of the paper. S.C. performed some of the additional experiments described in the Supplementary Information, and carried out important parts of the analysis. S.C. also helped substantially with understanding the results of the mass dependence of the inflow process as well as with the writing of the final version of this Letter.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to L.M. (lmayer@physik.unizh.ch).